THE BOSE-EINSTEIN ATOM • NANOLASERS • DRUG-RESISTANT BACTERIA

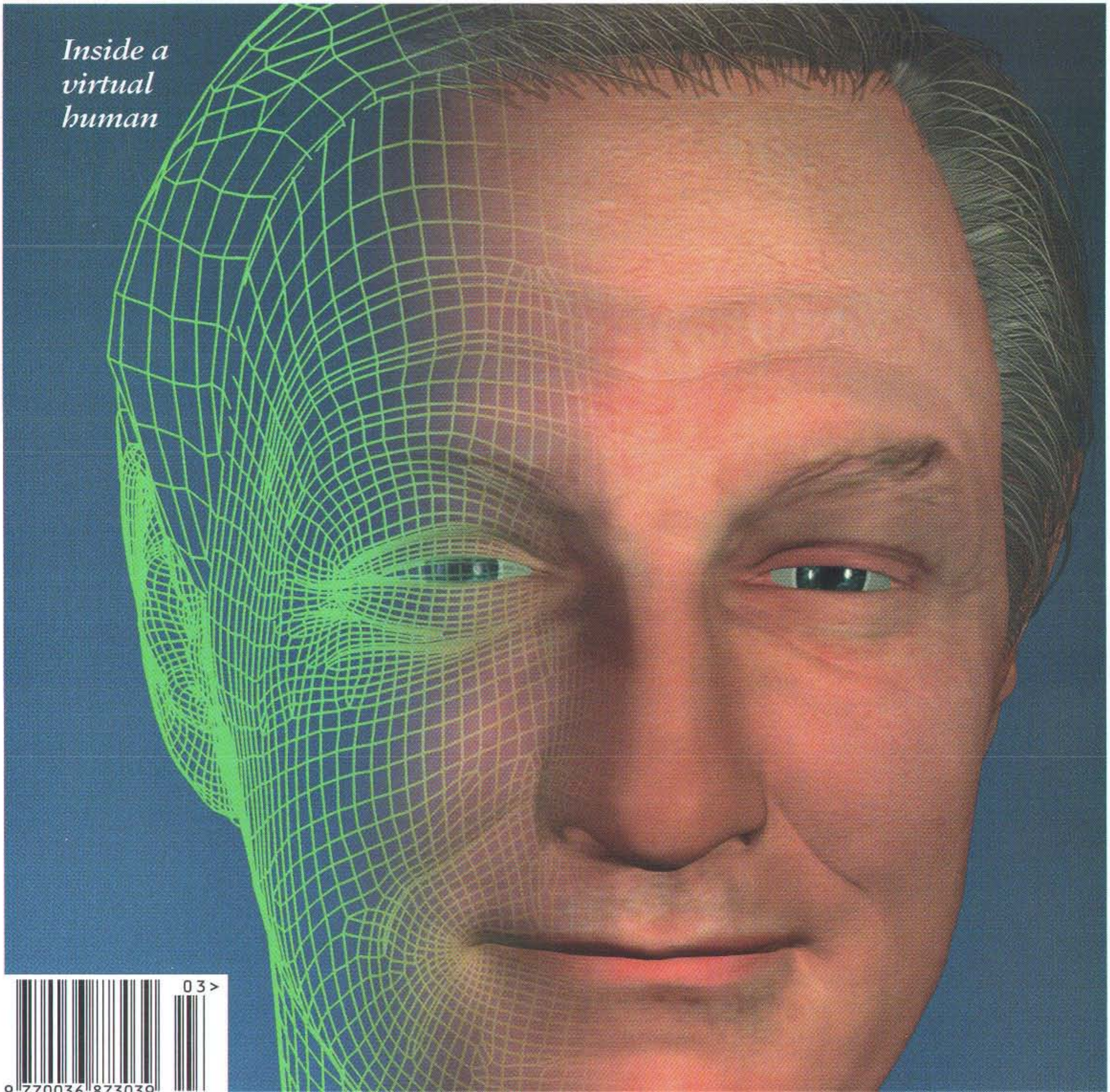
SCIENTIFIC AMERICAN

MARCH 1998 $4.95 U.K. £3.00

SPECIAL REPORT:

THE END OF CHEAP OIL

It's Coming Fast. But New Technologies Might Prevent an Energy Crisis

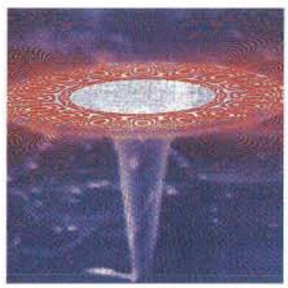



click here:
www.
sciam.
com
and go beyond the

SCIENTIFIC AMERICAN

March 1998 Volume 278 Number 3

SPECIAL REPORT:



Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111.

Periodicals postage paid at New York, N.Y., and at additional mailing offices. Canada Post International Publications Mail (Canadian Distribution) Sales Agreement No. 242764. Canadian BN No. 127387652RT; QST No. Q1015332537. Subscription rates: one year $34.97 (outside U.S. $47). Institutional price: one year $39.95 (outside U.S. $50.95). Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537. Reprints available: write Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111; fax: (212) 355-0408 or send e-mail to sacust@sciam.com **Subscription inquiries: U.S. and Canada (800) 333-1199; other (515) 247-7631.**

REVIEWS AND COMMENTARIES

About the Cover
This virtual replica of Alan Alda, the host of SCIENTIFIC AMERICAN *Frontiers,* has a texture map of his flesh over a digital wire frame of his facial structure. Animating such a construction realistically is a challenge. Image by Lamb & Company; photographic composition by Slim Films. Cover wrap photograph by Sam Ogden.

Visit the SCIENTIFIC AMERICAN Web site (http://www.sciam.com) for more information on articles and other on-line features.

FROM THE EDITORS

Screen Idols and Real Stars

Computer scientists (and maybe a few frustrated directors) have speculated about replacing human actors with digitally synthesized performers. Special effects have become so important to films, why not clear the set altogether? In theory, "synthespians" can do anything a script requires, from professing love to singing an aria to stomping on a skyscraper. They can combine the best features of a dozen mortals: care for a leading lady with the smile of Julia Roberts, the eyes of Isabella Rossellini and the cheekbones of Rita Hayworth? The current movie *Titanic* features computer-generated people moving on deck, but—sorry, kid, don't call us, we'll call you—their veneer of realism cracks under scrutiny. Still, it won't be long before somebody from central (processing unit) casting is ready for his or her close-up.

For the latest episode of SCIENTIFIC AMERICAN *Frontiers,* host Alan Alda lent his form and voice to an attempt to create a "Digital Alan." He was a fitting choice for this new medium, given that he is already a veteran of film, television and the stage. The process and results are described on pages 50 and 51 of this issue and on *Frontiers* (check your local listings for time and channel). Remarkable though Digital Alan is, I don't think he'll be stealing any roles from the real thing. Real talent is never obsolete: performance is all about expressing humanity.

DON DIXON

This month we announce the debut of a new sibling magazine. Four times a year *Scientific American Presents* will turn its attention to a single topic, with the depth of coverage that loyal readers will associate with this magazine's special issues.

The premiere issue, *Magnificent Cosmos,* reports on the many surprises emerging from astronomy. Leading authorities discuss planetary science, the sun, stellar evolution, the structure of the universe, dark matter and much more; every contribution, including classic articles that have been completely revised and updated for this volume, is fresh and rewarding.

Future issues of *Scientific American Presents* will consider women's health, the International Year of the Ocean, the riddle of intelligence—any and all areas of ongoing investigation. We hope that *Scientific American* and its new partner will jointly offer our readers (the most intellectually insatiable audience in known space) the comprehensiveness and rigor they desire and expect.

JOHN RENNIE, *Editor in Chief*
editors@sciam.com

SCIENTIFIC AMERICAN®
Established 1845

SCIENTIFIC AMERICAN, INC.
415 Madison Avenue • New York, NY 10017-1111
(212) 754-0550

PRINTED IN U.S.A.

LETTERS TO THE EDITORS

SCIENCE-BY-MAIL

In the November 1997 "Anti Gravity" column, Steve Mirsky reports on the persistent stereotyping of scientists in our culture ["The Big Picture," News and Analysis]. One need only look at popular culture to see why. Without exception, scientists and "smart kids" are portrayed in movies and on television as weird, unsocial and, of course, wearing glasses (and a bow tie if male).

Mirsky mentions Nerdkids trading cards as one of several efforts to counteract this trend, but the cards miss the mark by an obvious mile. One program not mentioned is Science-by-Mail, run by the Museum of Science in Boston. This program provides a way to connect kids with real scientists who act as mentors. Science-by-Mail should help change perceptions of what scientists are really like.

PETER OLOTKA
Centerville, Mass.

MUSEUM OF SCIENCE

***VISIT THE MUSEUM OF SCIENCE** on the World Wide Web at http://www.mos.org for information on the Science-by-Mail program.*

PLANTS AS DRUG FACTORIES

I read with much interest the news report by W. Wayt Gibbs on human antibodies produced by plants ["Plantibodies," News and Analysis, November 1997]. Unfortunately, the report neglected to point out a major problem that needs to be overcome before plants can be used for the production of pharmaceutically useful glycoproteins such as antibodies. The problem stems from the fact that the carbohydrate group attached to proteins by plants is often antigenic in humans and may be a strong allergen. Much has to be done to make plants the drug factory of the future.

NATHAN SHARON
Weizmann Institute of Science
Rehovot, Israel

Gibbs replies:

Sharon is correct: as I noted in the story, plant cells can stick the wrong carbohydrates onto a human antibody. This usually inactivates the plantibodies; sometimes it may cause an allergic reaction. As I reported, companies developing plantibodies say they have selected drugs that should work without any carbohydrates attached at all. Clinical trials that have begun recently will test that claim.

TRIGONOMETRY TEST

In the November 1997 article "Fermat's Last Stand," by Simon Singh and Kenneth A. Ribet, the authors present the equation "$\sin q = \sin(q + \pi)$." As it happens, I recently visited a high school trigonometry class where the instructor was at some pains to show that $\sin q = -(\sin q + \pi)$. The authors, apparently worried about making clear a complex proof, passed over this rather basic error.

MARK VAN NORMAN
Berkeley, Calif.

Editors' note:

We apologize for the error: the correct equation is $\sin q = \sin(q + 2\pi)$.

MISSION TO MERCURY

Robert M. Nelson's article on Mercury addressed a long-unfilled need in the area of planetary sciences ["Mercury: The Forgotten Planet," November 1997]. Ever since the possibility of ice on Mercury was announced, my imagination has worked overtime. I've got a not so rhetorical question: What would it take to get an international expedition to support the idea of a lander and rover with an ice-core sample return mission? The technology spin-offs to create a device capable of landing in sunlight, roving into darkness, obtaining the desired samples and returning them intact should be incentives for industry to participate. For the rest of us, it would be a firsthand look at the tree rings of the solar system.

DAVID B. LANGLOIS
Lafayette, La.

Nelson replies:

Langlois is quite correct. A deep-space-exploration mission to Mercury is long overdue. I understand that there are at least four teams planning to submit proposals this spring to the National Aeronautics and Space Administration for missions to Mercury. If approved, some could enter Mercury orbit as early as 2004; others may be as late as 2008.

WALKING ROBOT

I was pleased as punch to see a picture of my robot in the November 1997 issue ["'Please, No Double-Sticky Tape,'" by Marguerite Holloway, News and Analysis]. The caption under the picture read "Death on Wheels: Razor Back and others conform to technological correctness"—pretty funny, considering my robot is named "Pretty Hate Machine" (or "P.H.M.") and is actually a walking robot, not a rolling one. A better caption would have been "Death on Legs," but then again, it doesn't have quite the same ring.

CHRISTIAN CARLBERG
North Hollywood, Calif.

Letters to the editors should be sent by e-mail to editors@sciam.com or by post to Scientific American, 415 Madison Ave., New York, NY 10017-1111. Letters selected for publication may be edited for length and clarity.

ERRATUM

In the article "The Longest Suspension Bridge," by Satoshi Kashima and Makoto Kitagawa [December 1997], the Brooklyn Bridge was incorrectly identified as the world's first suspension bridge. It is the world's first such bridge to use steel wire for the cables. The design for suspension bridges was first developed in India in the 4th century A.D.

50, 100 AND 150 YEARS AGO

SCIENTIFIC AMERICAN

MARCH 1948

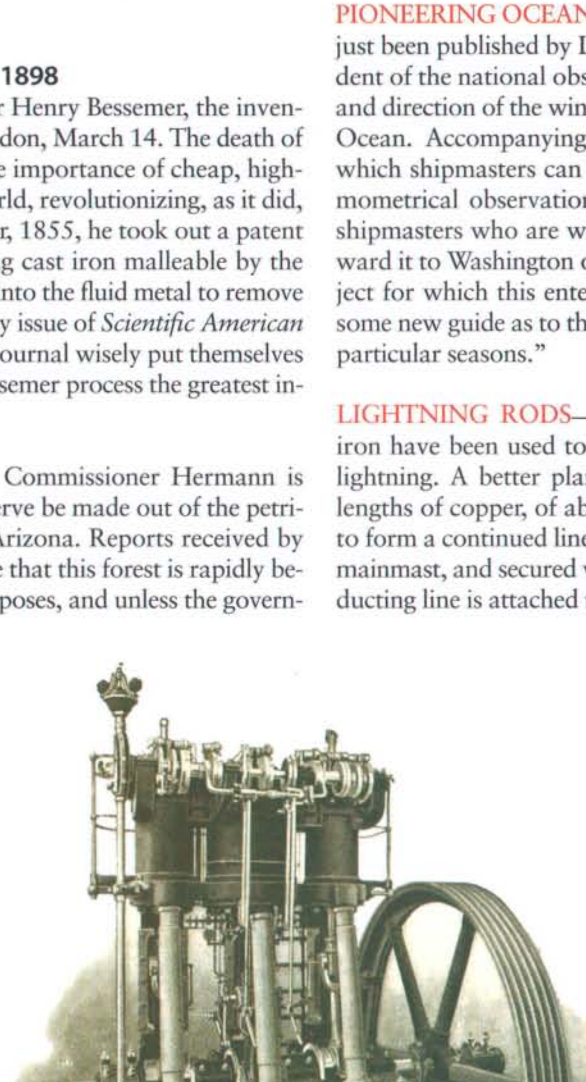
RADIO TAKES OVER—"While post-war radio has not been all it was quacked up to be, FM and Television are two significant developments on the horizon. Television is still in its early stages of development and is so expensive as to be in the luxury class, but FM is now coming into its own. One large-scale dealer is now advertising a portable model FM-AM receiver for $54.95 [about $400 in 1998 dollars], which brings this kind of radio reception down to the prices the average man can afford to pay."

MARCH 1898

STEEL AND PROGRESS—"Sir Henry Bessemer, the inventor and metallurgist, died in London, March 14. The death of this great man reminds us of the importance of cheap, high-quality Bessemer steel to the world, revolutionizing, as it did, many vast industries. In October, 1855, he took out a patent embodying his idea of rendering cast iron malleable by the introduction of atmospheric air into the fluid metal to remove carbon. In the fiftieth anniversary issue of *Scientific American* [July 1896], the readers of our journal wisely put themselves on record as considering the Bessemer process the greatest invention of the last fifty years."

PETRIFIED FOREST—"Land Commissioner Hermann is recommending that a forest reserve be made out of the petrified forest in Apache County, Arizona. Reports received by the Interior Department indicate that this forest is rapidly being used up for commercial purposes, and unless the government steps in to stop the despoilment, the whole forest, which is one of the greatest natural curiosities in the world, will disappear. A hotel is being built in Denver, all the walls of which are to be faced with the silicified wood taken from the forest, and all the tables for the hotel are also to be made of it." [*Editors' note: The Petrified Forest was proclaimed as a national monument in 1906.*]

DIESEL'S MOTOR—"An advance as important as the introduction of the internal combustion motor has been made by Mr. Rudolph Diesel, of Munich. The experiments which led to the construction of the present successful machine (*at right*) began in 1882. In the ordinary gas or oil engine, the charge within the cylinder is ignited by a jet, hot tube or electric spark. In the Diesel motor the temperature of ignition is secured by the compression of pure air. Air is compressed to a pressure of up to 600 pounds to the square inch and the fuel, kerosene, is injected gradually into the cylinder and is burnt steadily during the stroke of the piston."



The new Rudolph Diesel 150 horse power motor

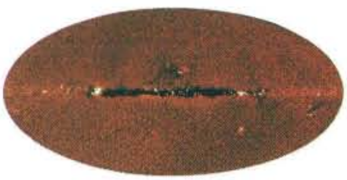
MARCH 1848

PIONEERING OCEANOGRAPHY—"A series of charts has just been published by Lieut. Matthew F. Maury, superintendent of the national observatory, designed to show the force and direction of the winds and currents of the North Atlantic Ocean. Accompanying these charts is an abstract log in which shipmasters can enter their daily run, currents, thermometrical observations, &c. The charts will be given to shipmasters who are willing to keep the above log and forward it to Washington on their return. The praiseworthy object for which this enterprise was undertaken is to provide some new guide as to the course which vessels should steer at particular seasons."

LIGHTNING RODS—"Chain conductors of copper and iron have been used to prevent ships from being struck by lightning. A better plan has been contrived. It consists of lengths of copper, of about four feet, rivetted together so as to form a continued line. This is inlaid at the after part of the mainmast, and secured with copper nails. In the hull, the conducting line is attached to the keelson. A square-rigged vessel was fitted with this apparatus, and a powerful electric discharge was communicated to the point of the main top gallant mast. The electric fluid passed along the conductor, and out of the vessel, without injuring any thing."

ELECTRIC TENSION—"A communication to the Paris Academy of Sciences from Monsieur Pallas suggests that the greater number of nervous affections are occasioned by the excessive influence of atmospheric or terrestrial electricity. He states that by adding glass feet to bedsteads and isolating them about eighteen inches from the wall, he has cured the patients sleeping upon them of a host of nervous affections."

NEWS AND ANALYSIS




IN FOCUS

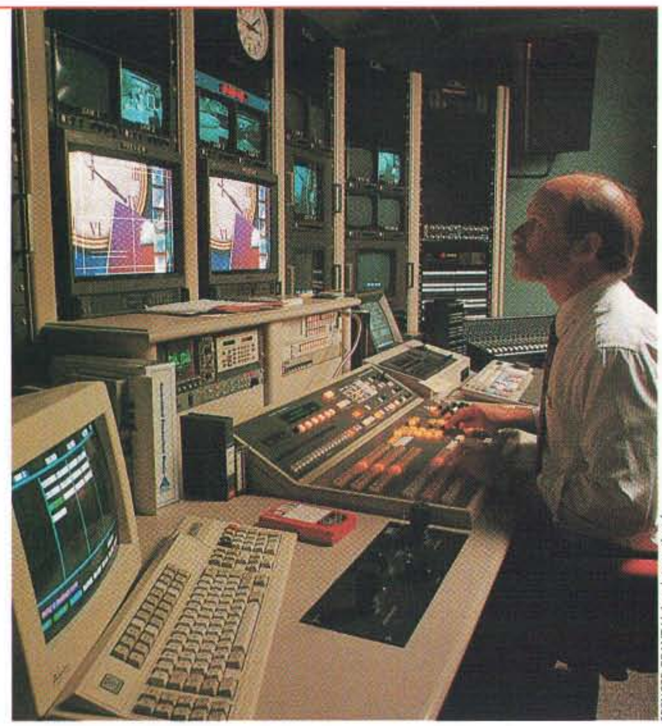
DIGITAL DILEMMA

The upcoming digital format for television lends itself to computer use, but disagreements about transmission standards could affect the melding of the two

Convergence—the industry jargon for the merger of television and the personal computer into one interactive appliance—is upon us. With the advent of digital television (DTV), the commingling seems inevitable: DTV's enormous bandwidth, or line capacity, permits both television reception and connection to the Internet, which the existing broadcast format cannot handle. Yet the form of the coming merger is still up in the air. Whether it will be TV-centric or accommodating to PC users depends on behind-the-scenes business decisions on transmission standards that will play out in the coming months.

DTV's rollout begins in November, only 18 months after the Federal Communications Commission lent each of the nation's TV stations a second channel based on pledges to broadcast some high-definition programming. Most of the new channels lie within the UHF band (numbers 14 to 59); existing channels are to be given back once the DTV transition is complete.

The huge capacity of DTV means that broadcasters can fill a channel with one crystal-clear, high-resolution program with six-channel sound and a wide screen—or put four or five standard-definition shows in the same space. Initially, four major network affiliates in the top 30 markets will begin digital broadcasting, offering a mix of standard-definition TV (SDTV) and high-definition TV (HDTV), probably for movies and sports (note that the "D" here does not stand for digital).



SPENCER GRANT *Liaison International*

TELEVISION BROADCASTING ***fundamentally changes in the U.S. in November, when digital transmissions start in the top 30 markets.***

The generous bit stream means that even an HDTV movie leaves room to transmit World Wide Web pages simultaneously. DTV will permit tierloads of customized news, music, sports, college courses, interactive games and catalogues. For this reason, the "D" in DTV for many stands as much for "datacasting" as it does for "digital." The worldwide wait over the telephone lines may finally come to an end.

But it's not going to be that straightforward. The FCC mandate of 1996, which allocated the digital airspace, specifically omitted the viewing format for DTV, leaving the issue to the

market to decide. Broadcasters want to transmit interlaced HDTV signals; computer makers prefer progressive scanning. The formats are quite different. In interlacing, the video camera creates one field of video that has even-numbered lines and then, in a second scan, creates a second field with odd-numbered lines (the present analog standard is 525 interlaced lines to a frame, scanned at a frame rate of about 30 hertz, and is called 525I). In progressive scanning, the video creates all the lines in order for each frame, as do computer displays, which require sharpness for text. Historically, interlace was one of the few ways to compress a TV signal. Nowadays the vertical and temporal resolution it produces can be accomplished with modern digital-compression systems. But after more than five decades of using the interlaced format for production and transmission, broadcasters have become accustomed to it. They also have substantial investments in the equipment that supports it. So when the FCC let the video-transmission format remain open to market forces, broadcast and computer camps went head-to-head.

The issue is not so much what's good about progressive but what's bad about interlace. "It's a roadblock on the way to convergence," says Alvy Ray Smith, a graphics fellow at Microsoft. "It accommodates only low-resolution text and graphics if you want to avoid flicker." The Web is full of text and graphics and hence inherently ill suited to interlace scanning.

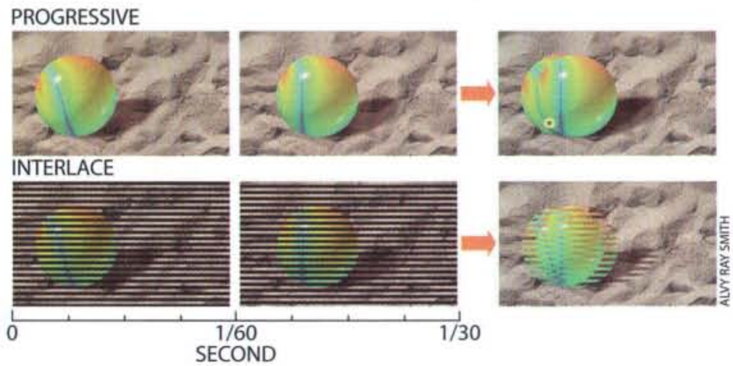
If interlace becomes the de facto standard for HDTV transmission, displaying the signals on a progressive scan monitor (a computer screen) is going to involve costly circuitry, keeping the market for PC-based TV small. "Viewers will need an expensive board to convert interlaced HDTV transmission to progressive. The board could easily cost $1,000," Smith says. "Even at a price, the de-interlacing will not be perfect and will result in a poorer image." Interlace also reduces the opportunity for datacasting, because it compresses less efficiently than progressive scanning does.

Tom McMahon, Microsoft's architect of digital television and video, says that even if TV people work in interlace inside the studio, what they should broadcast is another matter. "Once interlace is in the system, it's difficult to get it out. If there's any de-interlacing to be done, broadcasters should do it just prior to transmission to keep the receivers cheap." A microelectronics coalition of Intel, Microsoft, Compaq and Lucent Technologies, called the Digital TV Team, has worked to persuade broadcasters to initiate digital transmission using high-definition, 720-line progressive at 24 frames per second for material shot on film, and standard-definition, 480-line progressive for other content.

New York Times reporter Joel Brinkley, author of a lively, detailed history of the advent of HDTV, *Defining Vision: The Battle for the Future of Television,* says the point is already moot. "There is no standards battle," he claims. "The TV manufacturers have resolved it among themselves. I've spoken to all of them. They are going to build TVs that receive all 18 of the standards [for DTV] but display only two or three. They don't intend to support progressive except for 480-line format because of the cost. They won't support 720-line progressive, because the scanning frequencies would make TV too expensive by their definition. 1080I [interlace] will be the de facto standard for high definition."

"It's not up to TV manufacturers," Microsoft's Smith counters. "It's up to the broadcasters. Sure, right now, manufacturers are covering their bases by building a combination, but they'll be quick to build for whatever becomes the national signal."

Brinkley doesn't think the television industry will heed the computer manufacturers. "Certainly DTV offers the capability for incredibly fast downloads from the Internet," he says, "but broadcasters would have to devote a channel to Internet access, and I don't think they see this as a viable business plan right now. It may grow, but right now it's a niche market for broadcasters who reach every American home."



PROGRESSIVE AND INTERLACED SCANNING is simulated for a moving object. Progressive gives a full frame of information each instance; interlace scans every other line, filling in the remainder in the subsequent frame. After 1/30 second (two frames), the eye sees a blurred ball in progressive; in interlace, the image breaks up (human perception compensates somewhat, so the image is less objectionable in real video).

The conflict will not be settled until the broadcasters start writing checks for transmission equipment. "The networks have said they will announce their strategies early in 1998," Brinkley notes. "That will leave plenty of time for broadcasters to purchase the proper equipment." So far only CBS has gone on record for preferring 1080I.

Come what may, PCs can still converge with DTV—as long as the TV is standard rather than high definition. The do-nothing alternative for broadcasters is to air digital simulcasts of their current 525I analog programs. McMahon notes that such simulcasts are easily handled by PCs. "Interlace presents no difficulty so long as the program is in standard definition," he says. "The DTV receiver base we will begin to deploy in 1998 can receive these interlaced broadcasts, along with any data that might be transmitted concurrently."

PCs may indeed become a place where people watch standard-definition TV. Intel's Paul Meisner predicts that in the early years of DTV, TV manufacturers won't be able to grind out the new sets in great numbers. It will be easier to incorporate a digital SDTV receiver in a PC that will cost the consumer a few hundred dollars more, he says.

Terrestrial-TV-transmission debates may also take a new turn in coming months as cable companies mobilize to cash in on new digital services. Cable lines enter more than 90 million U.S. living rooms, and the number will jump with DTV because many homes will not receive over-the-air DTV without a large antenna. If the cable companies don't carry 1080I, the issue of interlace versus progressive may indeed become moot, to the benefit of PC users. *—Anne Eisenberg in New York City*

IN BRIEF

Iron Tooth
One of the oldest professions may be dentistry. Indeed, French researchers recently found a wrought-iron dental implant—in the upper right gum of a man's remains—in a Gallo-Roman necropolis dating to the first or second century A.D. Because the implant and the socket match perfectly and the iron and bone mesh, Eric Crubézy of Toulouse University and his colleagues conclude that the implant's maker used the original tooth as a model and hammered in the replacement. Ouch.

E-Test for Eyes
Computer screens cause tremendous eyestrain. To read fuzzy, pixelated letters, our eyes must refocus some 15,000

ANDY SACKS *Tony Stone Images*

to 20,000 times during the average workday, estimates Erik L. Nilsen of Lewis and Clark College. But a solution may be on the way. Nilsen has found that if you measure for eyeglass prescriptions using an on-screen eye test, as opposed to the printed E variety, you can minimize some of the eyestrain that screens cause.

Strange Stars
What happens once a massive star has collapsed into a stable collection of neutrons? Before 1984, astronomers thought nothing. Then Edward Witten of the Institute for Advanced Study in Princeton, N.J., proposed that they might further evolve into superdense wads of strange quarks—one of six types of quarks, which are the smallest constituents of matter. Now Vladimir Usov of the Weizmann Institute of Science in Israel has fully described just how neutron stars and strange stars would differ: strange stars would emit x-ray energy 10 to 100 times greater than that emitted by neutron stars; the x-rays would be fired off in one millisecond pulses; and strange stars would contain a small number of electrons and thus release high-energy gamma radiation. Based on these criteria, an object called 1E1740.7-2942, which is now believed to be a black hole, is a strong strange-star candidate.

More "In Brief" on page 12

SCIENCE AND THE CITIZEN

COSMOLOGY

GLOW IN THE DARK

A second cosmic background radiation permeates the sky

Modern theories of the universe begin with the simplest of observations: the night sky looks dark. The darkness implies that the universe is not infinitely old, as scientists once thought. If it were, starlight would already have seeped into all corners of space, and we would see a hot, uniform glow across the sky. This insight is known as Olbers's paradox, after the 19th-century German astronomer Wilhelm Olbers.

Some kinds of light, however, have had enough time to suffuse space. The famous cosmic microwave background radiation, considered to be the definitive proof of the big bang, fills the sky. Now astronomers say they have found a second, younger background. It is thought to be the first look at a previously unseen period of the universe—between the release of the microwave background and the formation of the earliest known galaxies, about a billion years later. "We're really completing the resolution of Olbers's paradox," said Princeton University astronomer Michael S. Vogeley, one of the researchers who announced their findings at the American Astronomical Society meeting in January.

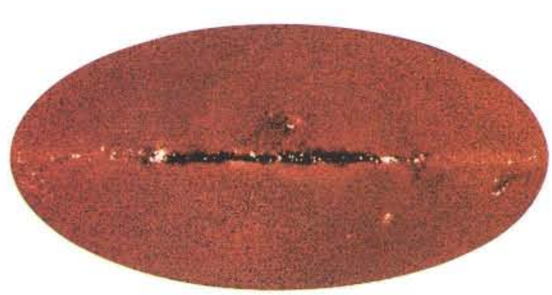
The greatest hoopla at the meeting concerned the far-infrared part of the background, first hypothesized in 1967 by R. Bruce Partridge of Haverford College and P. James E. Peebles of Princeton. Two effects turn primordial starlight into an infrared glow: the expansion of the universe, which stretches visible wavelengths of light into the infrared; and the presence of dust, which absorbs starlight, heats up and reradiates.

The background proved too dim to be seen by the Infrared Astronomical Satellite (IRAS) and other detectors previously. The decisive measurements were made by the Cosmic Background Explorer (COBE) satellite during 1989 and 1990, although it was not until 1996 that a group led by Jean-Loup Puget of the Institute of Spatial Astrophysics in Paris tentatively detected the background.

Now three teams have confirmed and extended Puget's findings. One, led by Dale J. Fixsen and Richard A. Shafer of the National Aeronautics and Space Administration Goddard Space Flight Center, used the same instrument on COBE—the Far Infrared Absolute Spectrometer (FIRAS)—that the French team did. Another, headed by Michael Hauser of the Space Telescope Science Institute and Eliahu Dwek of NASA Goddard, relied on COBE's Diffuse Infrared Background Experiment (DIRBE). A third team, led by David J. Schlegel of the University of Durham and Douglas P. Finkbeiner and Marc Davis of the University of California at Berkeley, combined DIRBE and FIRAS data.

No other COBE result demanded such arduous analysis. Starting with the total amount of observed infrared light, the researchers had to subtract the so-called zodiacal light produced by dust within our solar system and infrared light from stars and dust in the rest of our galaxy. They were left with a faint, nearly uniform glow that exceeded the inherent instrumental error.

Although the teams took different approaches, all arrived at nearly the same background intensity: 2.3 times as bright as the visible light in the universe, according to Hauser. The first implication is that the universe is filled with dust—much more dust than in the Milky Way and nearby galaxies. The second is that some unidentified source generates two thirds of the light in the cosmos.



COSMIC INFRARED BACKGROUND is revealed after light from the solar system and the galaxy is removed.

MICHAEL HAUSER *Space Telescope Science Institute*, COBE/DIRBE SCIENCE TEAM AND NASA OFFICE OF SPACE SCIENCE

"I don't think we know where this radiation is coming from," said Princeton astrophysicist David N. Spergel. "This emission could be coming from big galaxies; it could be coming from a class of small galaxies in relatively recent times."

To locate the source, a group directed by Puget and David L. Clements in Paris has started the first far-infrared search for distant galaxies, using the European Space Agency's Infrared Space Observatory (ISO). Through the Marano hole, a dust-free patch in the southern sky, they discovered 30 galaxies—10 times more than IRAS surveys had implied and exactly the number required to explain the infrared background. Unfortunately, ISO couldn't get a fix on the galaxies' positions. Analogous efforts by Vogeley and others have already explained a similar remnant glow in visible-light images by the Hubble Space Telescope.

How do these background measurements affect theories of how and when stars and galaxies formed? The current thinking is that once star formation began, it slowly accelerated, peaked when the universe was about 40 percent of its current age and has since declined 30-fold. But the unexpectedly bright background may indicate that star formation got going faster and remained frenetic for longer. If so, theorists might need to revisit the prevailing theory of galaxy formation, which posits clumps of so-called cold dark matter and agglomerations of small protogalaxies into progressively larger units. "It would cause real trouble for the cold-dark-matter model," Partridge said. "I think it's safe to say that we're seeing more energy than in all current models."

Besides identifying the source of the background, observers want to measure the glow at shorter wavelengths, determine how it has varied with the age of the universe and look for fluctuations. Upcoming missions such as the Far Infrared Space Telescope may prove crucial. Meanwhile the light-subtracting techniques may improve measurements of other phenomena, such as large-scale galaxy motions and the expansion of the universe. In short, scientists are encountering a new kind of Olbers's paradox. The night sky isn't dark; it's too bright. *—George Musser*

MUSICAL ACOUSTICS

UNSOUND REASONING

Are wind musicians loving tropical woods to death?

At a recent conference on music and human adaptation at Virginia Tech, physicist John W. Coltman demonstrated what he first described in the early 1970s. After asking the attendees to divert their eyes, he played the same tune twice on the flute. He then asked whether anyone heard any difference between the two performances. No one spoke up; the two were virtually indistinguishable.

Then Coltman revealed his trick. The first time he performed the tune, he played it on a simple side-blown flute made of lightweight cherry wood. The second time he used a flute of identical design, except for one detail: it was made of concrete.

To anyone schooled in the physics of wind instruments, Coltman's point is old news. Whether the air is set to vibrate by an edge tone as on the flute, by a reed as with the clarinet or by buzzing lips as with the French horn, the sound itself comes from the vibrating air column inside the instrument. This sound is produced through the end or through open tone holes, not by vibrations of the instrument's body, as is true of string instruments. Dozens of published reports, some dating back 100 years, converge toward the same general conclusion: so long as the walls are thick enough to remain rigid—about 0.4 millimeter for metals, two millimeters for woods—and the inside walls are smooth, the kind of material is, for the most part, immaterial.

But to many musicians, even a mountain of research remains unpersuasive. "We all know that wood flutes are much more dolce, much sweeter," says flutist Paula Robison. In contrast, "a gold flute sounds like an instrument made of gold. The silver flutes are much more perky."

The variation in timbre of wood and metal instruments stems from differences in acoustic dimensions brought about by the manufacturing process, not by the materials per se, says Robin Jakeways, a physicist at the University of Leeds. For example, holes in wood flutes are simply drilled in, whereas metal flutes have holes enclosed in a short length of pipe. Brian Holmes, a physicist at San Jose State University and a professional

In Brief, continued from page 10

Bypassing Bypass Surgery

Doctors at New York Hospital–Cornell Medical Center, led by Ronald G. Crystal, have begun testing a promising new gene therapy: By injecting inactivated virus particles containing vascular endothelial growth factor (VEG-F) genes directly into a patient's heart muscle, they hope to prompt the organ to sprout its own bypass around arterial obstructions. VEG-F encodes proteins that direct the development of new blood vessels. Similar gene-based treatments have failed because the gene products must be continually produced to have any therapeutic value. But animal studies indicate that VEG-F need be present for only a short time to promote growth.

B is for Banana

Neuroanatomists have long credited the planum temporale—an inch-long stretch of brain tissue—with controlling language. In support of this view, earlier studies found that in humans, this structure is typically larger on the left than on the right. Recently, though, researchers from Columbia University, the Mount Sinai School of Medicine and the National Institutes of Health have found the same asymmetry in chimps. There are several possible explanations: some common ancestor developed this size discrepancy; chimps and humans use the larger left planum temporale for some purpose other than language; or chimps in fact have a far more complex language system than previously imagined.

KENNAN WARD *Bruce Coleman Inc.*

Mmmm... Steak

Back in 1991, scientists identified the first apparent odor receptors—transmembrane proteins attached to nerve cells in the nasal cavity that bind to molecules floating in the air and set off a series of chemical reactions. They could not, however, pair any one receptor to a single scent. Now Stuart J. Firestein and his colleagues at Columbia University have done just that, matching a receptor in rat nerve cells to octanal, which smells like meat to humans. Co-investigator Darcy B. Kelley describes the finding as "a Rosetta stone for olfaction."

More "In Brief" on page 14

horn player, cites a study that found that plastic and metal clarinets had tone holes with much sharper edges than their wood counterparts. When these holes were rounded off, these clarinets sounded much more like wood ones.

Materials also differ in their ability to conduct heat and vibrations. "While those vibrations may not affect the sound significantly, they certainly affect how the instrumentalist interacts with the instrument," Holmes explains. After spending a premium for an instrument made of expensive material, it's only human to convince yourself that you must sound better. And, as flutist James Galway points out, the workmanship of an instrument made of $70,000 worth of platinum is likely to be of extraordinarily high quality. "People pick up my flute and say, 'This is better.' Of course it's better; it's like getting into a custom-built motorcar," he says.

Whatever the underlying reasons, the devotion of many musicians to rare or precious materials could help contribute to their extinction. *Dalbergia melanoxylem,* known as M'Pingo, grenadilla (African blackwood) and *D. nigra,* also called rosewood or palisander, are considered endangered, says Richard F. Fisher, a forest scientist at Texas A&M University. Grenadilla is the wood of choice for clarinets, oboes and, increasingly, wood flutes and piccolos; rosewood is a favorite for recorders.

Although the demand for fine musical instruments might seem too small to inspire a debilitating harvest of the rain forest, Fisher asserts otherwise. To get to the remote regions where these trees grow, harvesters must clear rivers or build roads. "In many of these areas there are so many landless peasants looking for a piece of land to farm that after you remove just the few trees you want, they go in and invade because now they have access," Fisher says. "They cut down the rest of the forest...and start to grow crops."

Fisher adds that these tropical species are extremely difficult to raise on plantations. They take 60 years or more to reach maturity and tend to grow poorly when raised clustered together in stands, as their key defense against predation is being scarce in the forest.

Indeed, an instrument maker in Libertyville, Ill., Boosey and Hawkes, failed at replenishing M'Pingo trees, says François Kloc, a master craftsman there. To offer an alternative material, the company developed a "green" line of oboes and clarinets. These instruments are made of M'Pingo sawdust and a patented mixture of carbon fiber and epoxy glue that is heat-treated and placed in a press to give it the density of whole wood. This process enables the company to use all of the tree instead of only the prime 20 to 30 percent that was usable before. Old, damaged clarinets can also be recycled in a similar way to make new ones.

Whether such innovations will ultimately be widely accepted by music lovers remains to be seen. "Most musicians and many listeners believe without question that the material of which a wind instrument is made has a profound effect on its tone quality," Coltman remarks. "After 100 years, scientists have still convinced nobody."—*Karla Harby in Rockville Centre, N.Y.*

UPI/CORBIS-BETTMANN

FLUTIST JAMES GALWAY LEADS NEW YORK CITY MUSICIANS in 1986. The demand for exotic wind instruments may damage rain forests.

ANTI GRAVITY

Urine the Money

Time was that the only consequence of drugs in urine was the confiscation of Olympic medals. Now, however, researchers have coaxed lab mice to produce a valuable pharmaceutical agent and in a way that makes it easy to harvest and to purify. Thanks to science that definitely qualifies as being of the "gee whiz" variety, the mice produce the drug in their bladders and simply turn it over to interested parties when they urinate.

Transgenic animals that can produce pharmaceutical agents have been in the works for years, but the bioreactor organ of choice has been the mammary gland—useful drugs derived from animal milk are now in human clinical trials. The idea for trying the same with urine started when Tung-Tien Sun of New York University Medical School published a paper in 1995 describing genes for proteins, called uroplakins, that get expressed only in the bladder. These uroplakins mesh together and probably have a role in maintaining a tidy lining, a highly desirable feature in a bladder.

MICHAEL CRAWFORD

Kenneth D. Bondioli of the U.S. Department of Agriculture's Gene Evaluation and Mapping Laboratory read Sun's paper and realized that it might be worth trying to tack something useful onto a uroplakin gene. At this point, David E. Kerr and Robert J. Wall, also at the USDA's gene lab, started shuffling genes. The aim was to get a useful human gene to hitch a ride on a uroplakin gene and find acceptance in the chromosomes of a fertilized egg. If they could pull that off, they could create a transgenic animal that produced the human gene's product only in the bladder—mixed up, of course, with the rest of the micturation.

So it came to pass that the research team created transgenic mice, with the gene for human growth hormone riding the uroplakin gene appropriately designated UP2. And indeed, when this mouse tinkles, it leaves human growth hormone in the cup. (Actually, it does its business on the benchtop, which—take note, new parents with nice furniture—the researchers covered with Saran Wrap for easy collection.) Any commercial application for the bladder bioreactor would involve the creation of larger transgenic animals, such as cows, able to produce urine in buckets rather than thimbles.

For this feasibility study, mice and human growth hormone make for a handy test system. "It was a two-pronged choice," says Wall, leader of the USDA group, who managed not to leak this work to the press prior to its publication in the January *Nature Biotechnology.* True, the molecule does have commercial application to treat dwarfism and potentially to enhance muscle strength in the elderly. More important for this first run, growth hormone gives itself away if any of it gets produced other than in the bladder—what transgenic researchers really do call "leaky expression." Specifically, you get really big mice, easy to weed out via visual inspection.

Harvesting drugs from urine is less oddball than it may appear. The widely prescribed drug Premarin, a type of estrogen, is collected from horse urine. And gonadotropins, used to enhance ovulation, come from the urine of human females.

Urine has some distinct advantages over milk as a vehicle for pharmaceuticals. Number one, as it were, urine contains few proteins naturally, so purifying the product should be easier than purifying milk, with its complex protein mix. Moreover, animals have to reach maturity to produce milk, whereas they start making urine from birth. Finally, all animals, male and female, urinate. So don't be surprised if, down the road, pharmaceuticals derived from urine make a big splash. *—Steve Mirsky*

***In Brief,** continued from page 12*

Universal Expansion

New results presented at the American Astronomical Society meeting in January confirm earlier analyses that the fate of the universe is to expand forever and at an ever increasing rate. Astrophysicists from Princeton University, Yale University and the Supernova Cosmology Project, an international research program based at Lawrence Berkeley National Laboratory, formed their conclusion after analyzing numerous supernovae. The results push the age of the universe back to as much as 15 billion years, which would solve the "age paradox" (some stars are seemingly older than the universe). But the studies may also doom a popular theory: the early universe may never have gone through a period of rapid expansion, called inflation.

COURTESY OF C. CARILLI

Reducing Roadkill

The Florida Department of Transportation has granted $40,000 to Florida State University researchers Joseph A. Travis and D. Bruce Means to study the feasibility of building amphibian crossings under U.S. Highway 319. The road will soon be expanded to four lanes, and many fear that the added traffic may further endanger the gopher frog (*Rana capito*) and the striped newt (*Notophthalmus perstriatus*)—under consideration by the U.S. Fish and Wildlife Service for status as threatened species. In particular, Travis and Means hope to test out different types of culverts to find out which design the animals would be most likely to actually use. New England installed the first amphibian tunnels, but whether such crossings are successful is still unclear.

Prescribing for Dollars

Patient beware: The use of calcium-channel blockers for treating hypertension has stirred up quite a bit of controversy. Studies have shown that the medications can put individuals at a greater risk for heart attack. Still more troubling, a recent study by Allan S. Detsky of the University of Toronto and his colleagues found that physicians in favor of calcium-channel antagonists were far more likely than neutral or critical physicians to have financial relationships with pharmaceutical houses.

—Kristin Leutwyler

FIELD NOTES

YOU SEE BRAWNY; I SEE SCRAWNY

The pathological preoccupation with muscle building

I am happy with my body. I lift weights about three times a week and also skate, swim and run, and I make a show of eschewing the fatty, nutritionally bankrupt confections served at Scientific American's editorial meetings.

But I can't call myself a bodybuilder, because I don't have the requisite rocky ridges, lithoid lumps and spaghetti-veined bulges upon which to gaze lovingly in a mirror. And even if I had all those things, I'm no longer sure that I'd see them. Apparently, as some in the muscled class stare into the looking glass, they see not thermonuclear thews but rather skin and bones.

This dislocation between perception and reality goes by the name "muscle dysmorphia," and it is part of a larger group of disorders in which the afflicted fixate despairingly on a facial feature, body part or their entire bodies. It is the malady of the moment, thanks to a recent blitz of media coverage that included entries from the *New York Times*, *Muscle & Fitness* and the journal *Psychosomatics*. Harrison G. Pope, Jr., the psychiatrist who has done more than anyone else to uncover and describe the condition, says you can spot muscle dysmorphics by their "pathological preoccupation with their degree of muscularity." Pope, probably the only Harvard Medical School professor who can squat 400 pounds, pumps iron six days a week. He himself does not have muscle dysmorphia, he says. (But he did ask Scientific American not to disclose his height, weight or body-fat percentage.)

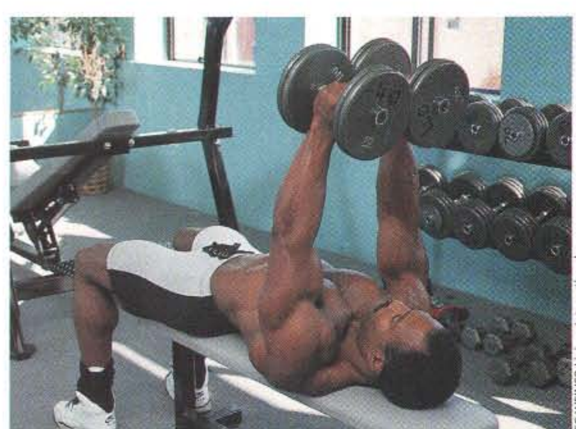
So how exactly do you pick out the pathological preoccupation with physique in a roomful of people staring at themselves in mirrors? You may have to leave the weight room and go door to door. Pope and some of his colleagues have found muscle dysmorphics with body weight well in excess of 200 pounds who were so ashamed of their puniness that they hardly ever left their homes, preferring instead to adhere to an indoor regimen of weightlifting in the basement, interrupted mainly for periodic protein consumption. This underground nature of the condition makes it virtually impossible to estimate how many people have it, Pope notes. Pope has also encountered dysmorphics who left high-powered jobs in law or business because their careers were taking away too much time from the gym.

Although use of anabolic steroids is not at all universal among muscle dysmorphics, Pope found "many who would persist in taking anabolic steroids or drugs for fat loss even if they were getting pronounced side effects." Such as? Well, beards and rich baritone voices in women, and high cholesterol, hard arteries and teeny testicles in men.

Intrigued, I began discussing the condition with my lifting buddies at the Vanderbilt YMCA in New York City. I found there is a little bit of muscle dysmorphia in almost everyone who hoists iron. "Don't we all suffer from that?" asked Chris, 39, a litigator who lifts six days a week and also boxes and runs. "It's just a matter of degree."

On a tour of New York weight rooms, I interviewed 13 men, ranging in weight from 185 to 290 pounds. All except one man thought they were not "big" or were not entirely satisfied with their size. Some used the word "small" to describe themselves.

At Diesel Gym on West 23rd, I spoke with Yves, 33, who is 5 feet, 10 inches tall and weighs 210 pounds. He lifts weights about 10 hours a week and says, "I'm happy, but I would like more size.



GREG WEINER *Liaison International*

PUMPING IRON can become an unhealthy obsession.

Always." He used to take steroids but swears he is now off the juice. Were the side effects too much? "No; it was very expensive," he explains.

At Dolphin on West 19th, I met Fabian, a six-foot-tall, 258-pound, 34-year-old professional wrestler who lifts five times a week. In Fabian's mind, Fabian is "an average Joe. I feel like I weigh 150 pounds. People look at me and say I'm big, and I take it as a compliment. But I see myself like I weigh 150 pounds."

The one content lifter I met was at Johnny Lats Gym, a 24-hour-a-day hard-core joint on East 17th. Kevin, a 27-year-old bodyguard, stands 6 feet, 6 inches tall and weighs 290 pounds. He is a former defensive end at Michigan State. He's big, isn't he? "Yeah, I have size," he agrees.

In the end, it was Duncan (5 feet, 11 inches, 190 pounds) who summed it up best. "You look at me and see what I have; I look at me and see what I want."

—Glenn Zorpette
(5′ 10 1/2″, 167 lbs., 7% body fat)

BIOCHEMISTRY

CATCHING THE RAYS

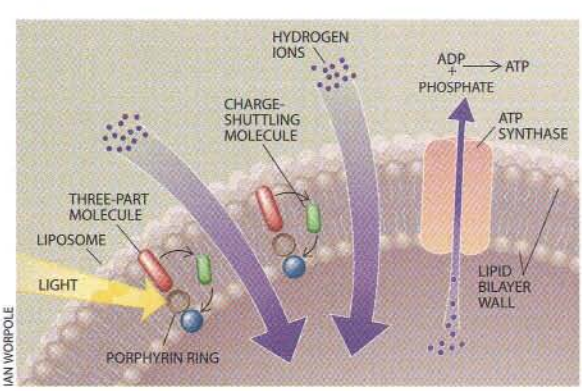
Researchers get a photosynthetic process in artificial cells

Essentially all of life is built from chemicals assembled with energy from the sun—directly through photosynthesis or indirectly via plant-derived fuel. Researchers at Arizona State University now believe they have succeeded in creating microscopic, solar-biological power plants that operate by a process modeled on photosynthesis. The investigators accomplished the long-sought feat in cell-like structures using only synthetic chemicals.

The work in Arizona employs water suspensions of liposomes—microscopic, hollow spheres with double-layered lipid walls. Liposome walls are in many ways like cell membranes, and by adding specific complex compounds to those walls, researchers, including Gali Steinberg-Yfrach, Ana L. Moore, Devens Gust and Thomas A. Moore, have made liposomes that act like the photosynthetic membranes inside plant cells. The minute vesicles first capture the energy from sunlight electrochemically; then they use it before it leaks away to generate ATP, a high-energy compound that cells employ as energy currency.

The new studies build on techniques that the Arizona group first described a year ago in *Nature*. Then, the scientists showed that incorporating two carefully selected compounds into liposomes gave them a rudimentary ability to capture light energy. One of these molecules was an elaborate, three-part affair featuring a porphyrin ring (a nitrogen-containing structure found in nature's photosynthetic molecule, chlorophyll). This three-part molecule spanned the liposome walls just as chlorophyll-containing proteins span photosynthetic membranes in plant cells. The second molecule moved around within the lipid walls. The two molecules functioned as a team. When illuminated, they joined in a cyclical reaction whose net effect was to drive hydrogen ions from outside the liposomes into the interior.

The Arizona researchers have now added a third molecule to the liposome walls that makes this laboratory curiosity potentially useful. The chemical is an enzyme called ATP synthase, which spans the liposome wall and returns the excess of hydrogen ions from inside back to the outside. This reverse migration releases energy that the ATP synthase exploits to generate ATP from its constituent parts, ADP and phosphate. According to Thomas Moore, who has described the new research at scientific meetings, the "proton motive force" created by excess hydrogen ions in the liposomes apparently generates ATP in "biologically relevant levels."



***LIGHT-CAPTURING MOLECULES** drive hydrogen ions into a liposome; when the ions leave, they power an enzyme that generates ATP.*

SCIENTIFIC AMERICAN

CORRESPONDENCE

Reprints: $4.00 each (minimum order, 10 copies) prepaid. Articles published within 3 months of current issue available. Write Reprint Dept., Scientific American, 415 Madison Ave., New York, NY 10017-1111.

Back issues: $9.95 each ($12.95 outside U.S.) prepaid. Most numbers available. Credit card (Mastercard/Visa) orders accepted. To order, fax (212) 355-0408.

Index of articles since 1948 available in electronic format. Write SciDex®, Scientific American Selections, P.O. Box 11314, Des Moines, IA 50340-1314, or call (800) 777-0444. E-mail: info@sciam.com

Photocopying rights are hereby granted by Scientific American, Inc., to libraries and others registered with the Copyright Clearance Center (CCC) to photocopy articles in this issue of Scientific American for the fee of $3.50 per copy of each article plus $0.50 per page. Such clearance does not extend to the photocopying of articles for promotion or other commercial purposes. Correspondence and payment should be addressed to Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Specify CCC Reference Number ISSN 0036-8733 /96. $3.50 + 0.50.

Editorial correspondence should be addressed to The Editors, Scientific American, 415 Madison Ave., New York, NY 10017-1111. Unsolicited manuscripts are submitted at the authors' risk and will not be returned unless accompanied by a stamped, self-addressed envelope. E-mail: editors@sciam.com

Advertising correspondence should be addressed to Advertising Manager, Scientific American, 415 Madison Ave., New York, NY 10017-1111, or fax (212) 754-1138. E-mail: advertising@sciam.com

Subscription correspondence should be addressed to Subscription Manager, Scientific American, P.O. Box 3187, Harlan, IA 51537. The date of the last issue of your subscription appears on each month's mailing label. For change of address, please notify us at least four weeks in advance. Please send your old address (mailing label, if possible) and your new address. We occasionally make subscribers' names available to reputable companies whose products and services we think will interest you. If you would like your name excluded from these mailings, please send your request with your mailing label to us at the Harlan, IA address. E-mail: customerservice@sciam.com

Visit our Web site at http://www.sciam.com/

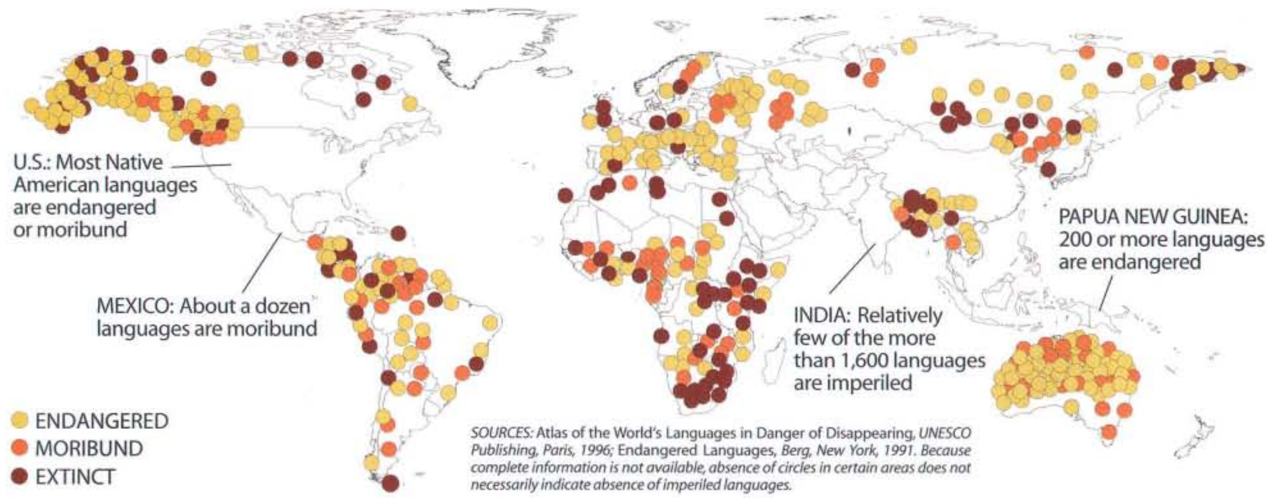
Moore says he and his colleagues demonstrate this with liposomes suspended in a solution of luciferase, which glows in the presence of ATP. When the solution is illuminated by a red laser beam, it emits the characteristic yellow light of luciferase activity, Moore describes. The work has been accepted for publication by *Nature*.

Moore, Gust and their associates plan next to couple other biological processes to their artificial biosolar cells, such as the molecular motors that rotate bacterial flagella and an enzyme that makes a biochemical known as NADH. If, as seems plausible, the researchers can make both NADH and ATP inside liposomes by supplying light, Moore speculates they should be able to drive a variety of biosynthetic and other reactions normally found only in living cells, such as the reactions used to create complex pharmaceuticals. Artificial biosolar systems "may be useful for powering small man-made machines that couple light energy to biosynthesis, transport, signaling and mechanical processes," Gust speculates. Moore, for his part, admits that the accomplishment "seems far-fetched." Just 10 years ago, he says, "I would not have believed this would be possible."

—*Tim Beardsley in Washington, D.C.*

BY THE NUMBERS

Languages, Disappearing and Dead



SOURCES: Atlas of the World's Languages in Danger of Disappearing, *UNESCO Publishing, Paris, 1996;* Endangered Languages, *Berg, New York, 1991. Because complete information is not available, absence of circles in certain areas does not necessarily indicate absence of imperiled languages.*

The death of languages has been repeated many times in history. Localized disasters such as great floods or warfare have played a part, but in the modern era the spread of Europeans and their diseases has greatly accelerated the destruction. Local languages may be overpowered by a metropolitan language, thus increasing the pressure to neglect the ancestral tongue in favor of the new one, which is seen as the key to prospering in the dominant culture. Children may be forbidden to use their mother tongue in the classroom, as has occurred to many groups, including the Welsh and Aboriginal Australians. Speakers of minority languages have been forcibly relocated and combined with speakers of other languages, as happened when Africans were brought to the Americas as slaves. Practices such as these have made Native American languages the most imperiled of any on the earth.

The death of a language is not only a tragedy for those directly involved but also an irretrievable cultural loss for the world. Through language, each culture expresses a unique worldview. Thus, any effort to preserve linguistic variety implies a deep respect for the positive values of other cultures. For these reasons, the United Nations Educational, Scientific and Cultural Organization (UNESCO) has taken an interest in the preservation of endangered languages and in 1996 published the *Atlas of the World's Languages in Danger of Disappearing,* which is the primary source of the map depicted here.

In addition to languages known to have become extinct in the past 400 years, the map shows two categories of imperilment: endangered, meaning those in which most children no longer learn the language and in which the youngest speakers are approaching middle age, and moribund, referring to languages spoken only by the elderly. The map is incomplete, for it is impractical to study all endangered languages, particularly those in remote areas. But the point is that every region, including Europe itself, is prone to language disappearance. Languages such as Norn (once spoken in the Shetland Islands) and Manx (Isle of Man) have succumbed to English, whereas in France, Breton and Provençal are seriously endangered.

To save the world's languages, linguists are following a twofold approach: for moribund languages, they attempt to preserve vocabulary, grammar, sounds and traditions so that scholars and descendants can learn them later. Many linguists—such as Stephen A. Wurm of Australia National University, the editor of the UNESCO atlas—believe moribund languages should be given priority because they are in imminent danger.

In the case of endangered languages, linguists can give advice on language maintenance and teach the language to young people. According to one estimate, about 3,000 languages—half of all those now spoken—are threatened with extinction. About half of these have been adequately studied, and several hundred more may be analyzed over the next decade. Given the low cost of doing a solid study of an imperiled language—often well under $100,000—the worldwide effort to preserve languages would seem to be a cost-effective cultural investment.

—*Rodger Doyle (rdoyle2@aol.com)*

RODGER DOYLE

Undressing the Emperor

Physicist and Social Text *prankster Alan Sokal fires another salvo at thinkers in the humanities*

"I'm not modest," concedes Alan Sokal, professor of physics at New York University, with a grin. "My parody is hilarious!" Two years ago his nonsensical article on the "hermeneutics of quantum gravity" appeared in the journal *Social Text,* only to be exposed by the author as a hoax. The Sokal "affair"—his detractors prefer "stunt"—highlighted misuses of scientific ideas by nonscientists, provoking front-page articles in the *New York Times,* the *International Herald Tribune,* the *London Observer* and *Le Monde* and a series of debates on university campuses. Sokal has now upped the ante by publishing, with physicist Jean Bricmont of Catholic University of Louvain in Belgium, a dissection of what he calls "sloppy thinking" on the part of postmodernists, social constructivists, cognitive relativists and sundry other "-ists."

The book, *Imposteures Intellectuelles,* is in French and primarily targets French thinkers, besides some English and American ones. (An English version is to appear later this year.) It made the best-seller lists in France—"French people take their intellectuals seriously," Sokal explains—and outraged his opponents. One more round of artillery had been discharged in the battle between scientists and their critics.

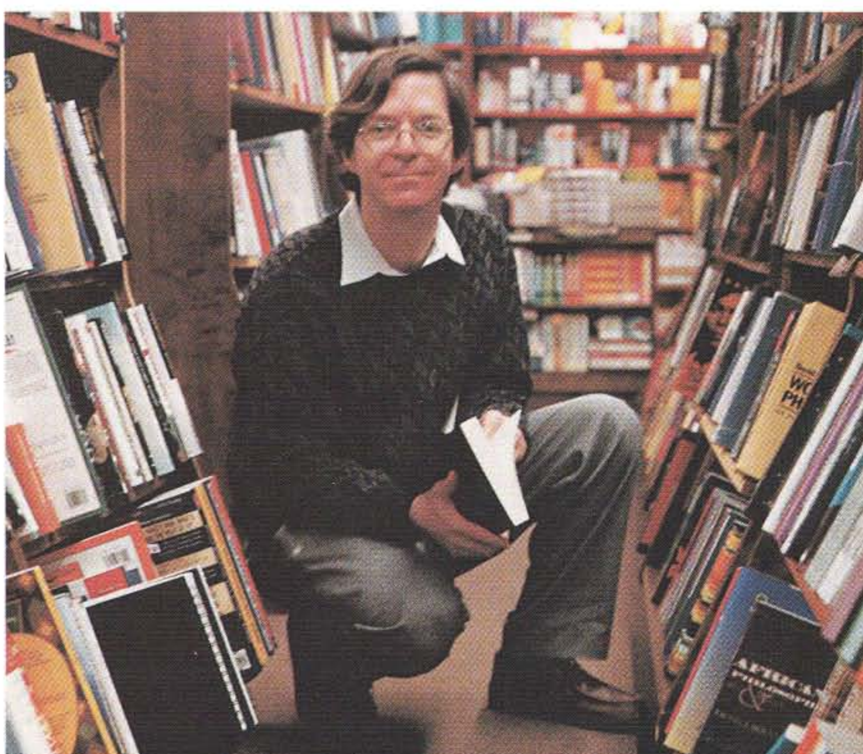
The floor of Sokal's office is strewn with papers, one side devoted to physics and the other to his newfound passion: defending the "scientific worldview." The two halves merge in the middle. I pick my way over the chaos to the worn sofa at the room's end. But Sokal jumps to his computer so he can refer to his Web site, and I have no choice but to return, sit at his feet on a hard book, and listen.

"I'm distressed by sloppy thinking," Sokal declares. "I'm distressed by especially the proliferation of sloppy thinking which confuses valid ideas with invalid ones." What most concerns him—apart from misappropriations of "the uncertainty principle," "Gödel's theorem," "chaos" and other terms from the physical sciences—is relativism or social constructivism as applied to science. "Roughly speaking," he explains, stressing particular words as though lecturing, "the idea is that there are no objective truths either in the social sciences but *even* in the natural sciences; that's what's most shocking—somehow that the validity of any statement is *relative* to the individual making it or relative to the social groups or culture to which that individual belongs."

He articulates clearly, stringing together astonishingly intricate sentences. I ask him how many scholars actually take such a position. "Well, very few people would say it in so many words, so explicitly and so precisely. But they say *vague things* that come down to that, if taken seriously. If you press them on that, they might come up with 'Oh, what I *really* meant is not the radical thing it seems to mean, but what I really meant was blah blah blah,' where blah blah blah is something that's not only *not* radical at all, it's true and trivial."

Thomas F. Gieryn, a social scientist at Indiana University, argues that the problem with the so-called science wars is that "neither side sees itself in the way the other side sees them." In 1994 mathematician Norman J. Levitt and biologist Paul R. Gross published *Higher Superstition: The Academic Left and Its Quarrels with Science,* a tract that assailed a motley collection of figures—social scientists, relativists, feminists, ecoradicals and others—as being hostile to science. The polished but occasionally pugilistic text (the authors express the hope that "the painful bolus of postmodernism will pass through the costive bowels of academic life sooner rather than later") distressed its subjects, many of whom claim their ideas were misrepresented.

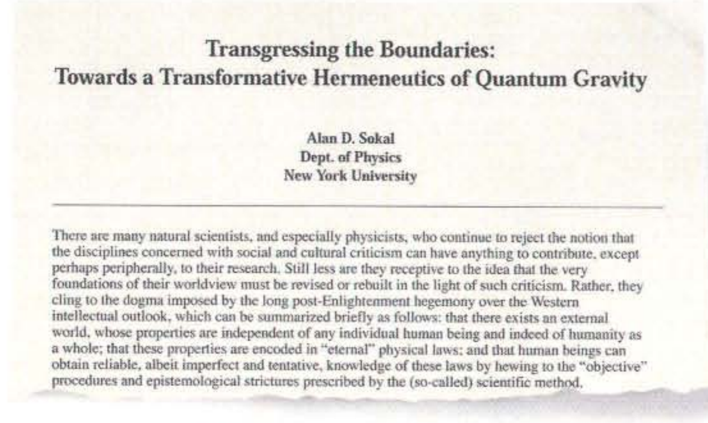
The book, however, inspired Sokal to visit the library and cull a dossier of delectable quotes, around which he crafted the hoax. "It took me a lot of writing and rewriting and rewriting before



MARK HAVEN

PUBLIC ATTENTION TO THE "SCIENCE WARS" resulted from Alan Sokal and his elaborate hoax.

the article reached the desired level of unclarity," he chuckles. After he revealed the hoax, Sokal recalls, he received a lot of e-mail from "people [in humanities and the social sciences] saying, 'Thank you, thank you, we've been saying this for years, and nobody listened to us; it took an outsider to come in and say that our local emperor is naked.'"



Transgressing the Boundaries:
Towards a Transformative Hermeneutics of Quantum Gravity

Alan D. Sokal
Dept. of Physics
New York University

There are many natural scientists, and especially physicists, who continue to reject the notion that the disciplines concerned with social and cultural criticism can have anything to contribute, except perhaps peripherally, to their research. Still less are they receptive to the idea that the very foundations of their worldview must be revised or rebuilt in the light of such criticism. Rather, they cling to the dogma imposed by the long post-Enlightenment hegemony over the Western intellectual outlook, which can be summarized briefly as follows: that there exists an external world, whose properties are independent of any individual human being and indeed of humanity as a whole; that these properties are encoded in "eternal" physical laws; and that human beings can obtain reliable, albeit imperfect and tentative, knowledge of these laws by hewing to the "objective" procedures and epistemological strictures prescribed by the (so-called) scientific method.

COURTESY OF ALAN SOKAL

SOKAL'S TONGUE-IN-CHEEK ARTICLE appeared in the "Science Wars" special issue of Social Text *(1996).*

Sokal moves incessantly within the small space defined by me and the sea of papers, jumping to his feet, bouncing on his toes, sitting down, pulling his legs up or swinging them out, perpetually propelled by the force of his convictions. His unconcealed glee at his achievement, combined with the restlessness, brings to mind a schoolboy who has caught his math teacher in a stupid mistake.

His objective, Sokal says, was not really to defend science. What upset him was that the relativism seemed to be emanating, in part, from the political left. "And," he protests, "I *too* consider myself politically on the left." Left-leaning academics, he believes, were dissipating their energies in pointless pontifications. "We need to develop an analysis of society which would be more convincing to our fellow citizens than the analyses they would read in *Newsweek* and the *New York Times*. And to do that we need good arguments. We can't afford sloppy thinking."

Despite his political motivations, however, Sokal's detractors see him primarily as a player in the science wars, and an aggressive one. I was repeatedly warned while researching this profile that anyone who steps into the war gets burned. The book review editor of *Science*, I was told, had been forced out because of a negative review she had published on *The Flight from Science and Reason*, a compilation of essays edited by Gross, Levitt and Martin W. Lewis. To be sure, the reviewer had resorted to personal gibes, but the *Chronicle of Higher Education* pointed to Levitt as having organized a campaign.

Levitt says that although he wrote a letter criticizing the quality of book reviews, accompanied by a note signed by several others, "There were lots of other letters [to *Science*] I had nothing to do with." Monica M. Bradford, *Science*'s managing editor, says the impending reorganization of the department had more to do with the editor's decision to retire.

Another oft-cited "casualty" is a faculty position in the sociology of science that went unfilled at the Institute for Advanced Study in Princeton, N.J. "Science wars had everything to do with it," asserts Clifford Geertz, who heads the school of social sciences. On two occasions, he says, the candidate of choice could not be appointed, because researchers within and without the institute mounted a campaign. "In the end, we decided we didn't have control over appointments," Geertz states, and the school returned the $500,000 for the position to the Henry Luce Foundation.

Although the initial candidate, Bruno Latour of the School of Mines in Paris, has made unflattering statements about the practice of science—which Sokal loves to quote—the last one, M. Norton Wise of Princeton University, has doctorates in both physics and history and sees himself as a mediator in the science wars. But he had an exchange with noted physicist Steven Weinberg of the University of Texas at Austin over the Sokal affair, a factor that Geertz mentions as having made Wise a target. "What I find most reprehensible about [the science wars]," Geertz laments, "is the political level to which it has sunk."

To his credit, Sokal was not implicated in either of these episodes. "I don't see this as a war," he protests. "I see this as an intellectual discussion." But his relentless ridicule frays the few and fragile ties that remain between scientists and humanists.

Hugh Gusterson, an anthropologist at the Massachusetts Institute of Technology, charges that Levitt, Gross and Sokal repeatedly confuse friendly critics of science with its enemies. "What they try to portray as antiscience thought," he adds, "is often a reflection on subtleties and complexities of the scientific method." Gieryn believes science and technology studies are a scapegoat: "We are being blamed for the failure of scientists to sell themselves politically." Scientists say, in turn, that humanists resent the special status enjoyed by scientists as custodians of objective knowledge; such "physics envy" leads to their attacking objectivity itself.

Thumbing through literature in science studies, I find myself in over my head. Most of the time I cannot figure out what is being said, let alone who is being critiqued. But in *The Golem: What Everybody Should Know about Science*, by sociologists Harry M. Collins of Southampton University in England and Trevor J. Pinch of Cornell University, I chance upon an essay on cold fusion. The writers suggest, without quite saying so, that cold fusion was stomped out because its proponents were politically weak compared with the powerful nuclear physicists and their vested interest in hot-fusion budgets.

As it happens, I was a nuclear physicist at the time news of cold fusion broke; now, in my head, I found myself protesting that most of us would have been thrilled if it had worked. The hostility emanating from the piece upset me, giving me a sense of how emotions on both sides of the war had come to run so high.

"My bottom line would be," Gusterson states, "that we need less name-calling and a more nuanced discussion that doesn't caricature the other side." In the pages of *Physics Today*, physicist N. David Mermin of Cornell has been engaging Collins and Pinch in just that, a quiet and uncommonly civil debate. And in July 1997 these and other scholars met in a modest "peace conference" at Southampton University.

The hoax had at least one positive effect. Out on a date with an Italian archaeologist, Sokal handed her the unpublished draft of his essay. Going home, she read it with increasing bewilderment, ultimately realizing that it was a spoof. "She was one of only two nonscientists who figured out it was a joke," he relates proudly. "Not the only reason I married her, of course."

—Madhusree Mukerjee

PUBLIC HEALTH

CLOSING THE BOOK

Are power-line fields a dead issue?

Last July, Edward W. Campion, a deputy editor at the *New England Journal of Medicine*, made a plea to kill studies that seek ties between power-line electromagnetic fields and cancer. "The 18 years of research have produced considerable paranoia, but little insight and no prevention. It is time to stop wasting our research resources," Campion wrote.

The editorial accompanied a report in the journal of a large epidemiological study by the National Cancer Institute that showed "little evidence" that magnetic fields from high-voltage lines, household wiring or appliances can increase the risk of childhood leukemia. It also followed by eight months a National Research Council review of hundreds of studies that indicated that "the current body of evidence does not show that exposure to these fields presents a human health hazard."

The controversy over electromagnetic fields (EMFs) provides a look at what happens to a science issue in the media spotlight when researchers repeatedly fail to find hard evidence that bears out public fears. Current U.S. research plans—motivated by both the lack of definitive findings and budget concerns—parallel Campion's urgings closely. As of the end of 1998, no federal program will be devoted to EMF research. The decision marks the finish of the world's largest effort, some $55 million in funding supplied during the past five years through the Department of Energy and the National Institute of Environmental Health Sciences. Another major funding source—the Electric Power Research Institute, the research arm of the utility industry—has reduced its contribution from a peak of $14 million in 1993 to roughly $6.5 million this year.

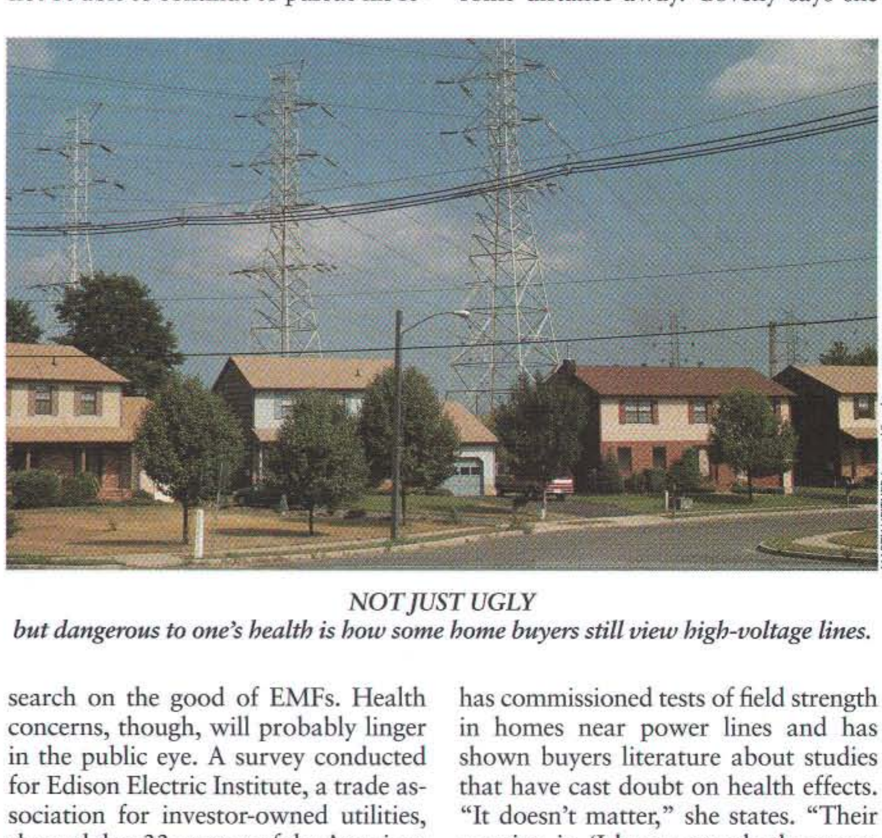
Researchers and activists who for years have pursued health issues related to EMFs lament the disappearance of support, pointing to unanswered questions that might possibly link residential fields to childhood leukemia, breast cancer and Alzheimer's disease, among other maladies. The NRC report did recommend some additional research—and investigators have begun to develop a rationale for continuing their labors.

One scientist has taken the novel approach of suggesting that some funding be devoted to examining the health benefits of EMFs. Theodore A. Litovitz, a biophysicist at Catholic University, who previously designed electric hair dryers and other devices that purport to block the effects of magnetic fields, has conducted unpublished studies that show that household fields can produce stress proteins in chick embryos that might protect against heart attacks. "I'm hopeful—we're beginning an era here where we have a much better reason for doing this research," Litovitz asserts.

In the current climate, Litovitz may not be able to continue to pursue his research on the good of EMFs. Health concerns, though, will probably linger in the public eye. A survey conducted for Edison Electric Institute, a trade association for investor-owned utilities, showed that 33 percent of the American public in late 1997 viewed EMFs as a serious health threat, up 8 percent from a year earlier but below the 41 percent figure registered in a poll taken in 1993.

Distress about EMFs has also had an effect on the U.S. economy. One of the few estimates ever made of their impact came in a 1992 article in *Science*. It suggested that concerns about EMFs cost the economy more than $1 billion a year, primarily because of modifications to new power lines to reduce exposure. Some states continue a policy of "prudent avoidance," requiring utilities to take reasonable mitigation steps for new lines. Utilities can configure them in a way that minimizes the strength of magnetic fields or can have them placed away from homes or schools.

Home values can still be affected. Although the National Association of Realtors receives fewer inquiries on EMFs, real-estate agents continue to contend with buyer anxieties, a problem also inextricably entangled with a dislike of the aesthetics of utility towers. Susan Coveny, president of RE/MAX Prestige, a realty agency in Long Grove, Ill., says a home near a power line can sell for 20 percent less than a comparable house at some distance away. Coveny says she has commissioned tests of field strength in homes near power lines and has shown buyers literature about studies that have cast doubt on health effects. "It doesn't matter," she states. "Their reaction is, 'I know somebody near a power line who has brain cancer.'"



MARTY HEITNER *Impact Visuals*

NOT JUST UGLY
but dangerous to one's health is how some home buyers still view high-voltage lines.

At the same time, the ongoing lack of scientific proof has not gone entirely unheeded by the courts. Worries over EMFs have not created "the next asbestos," a predicted bonanza for tort lawyers: no lawsuit claiming damage to physical health from proximity to power lines has ever held up in court.

Alleging damage to property values

has proved more fruitful, because the perception of threat can have a negative effect on prices even without strong scientific evidence. Courts in several states have ruled that juries can consider EMF fears in compensating owners for a decrease in value of their remaining land when utilities take property to construct new power lines. But even property cases have become more difficult to litigate. The California Supreme Court in 1996 upheld the decision of a lower court that blocks most lawsuits against utilities that claim loss in property value because of concerns about powerline EMFs.

Apprehension about power lines may gradually abate. But the public may just transfer its anxiety to another part of the electromagnetic spectrum. Conflicting assertions about the risks of using cellular phones could help fuel a separate health scare. Thus, the book may never fully close. *—Gary Stix*

MATERIALS SCIENCE

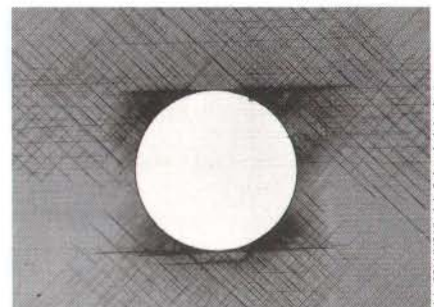
COMPOSITE SKETCH

Were composites to blame for recent aircraft accidents?

A pilot complained that the composite wing on a Stealth fighter plane seemed too flexible. His fears soon were realized: the $45-million aircraft shed that wing during an air show last September and plummeted to the ground. Not long after, singer John Denver died in the crash of a composite-construction homebuilt plane called a Long-EZ—the same type in which science writer James D. Gleick was gravely injured (and his son killed) when it crashed in New Jersey last December. Composites—which consist of a high-strength material, such as carbon fiber, that is impregnated with a resin and then cured—are lighter than the more traditional aluminum, are more easily formed and are often stronger. But compared with aluminum, which has been used in aircraft construction since the early 20th century, advanced composites are a recent technology, having been introduced in the 1970s. With relatively little experience with composites, how confident can engineers be about the lifespan and integrity of the material, which is being used more frequently in aircraft construction?

"Composites are so new we don't have a complete grasp of their durability," notes Chin-Teh Sun, an aeronautical and astronautical engineer at Purdue University who has extensively researched composites. What is known, he says, is that environmental factors such as ultraviolet light and moisture can change the materials' properties. And as is the case for any material, an excessive load can cause failure and damage. "But how long they will last nobody can tell you for sure," he adds.

There are some ways to approximate that information, however. At Virginia Polytechnic Institute and State University, Ken Reifsnider developed MRLife, a performance simulation code designed to predict the remaining strength and life of composite materials. Among other tasks, it analyzes a composite's molecular chemistry and the chemical and thermal details of its manufacture and processing. Combined with studies of the composite while it is in motion, MRLife estimates the rate at which a composite changes over time. "[The code] answers the question, 'Given these conditions and history and materials system, how long will it last and how strong is it after some periods of time?'" Reifsnider explains. Thus far, he adds, the code says composites will hold up better than expected.



SHEILA L. COLLINS *Virginia Polytechnic Institute and State University*

CRACKS IN A COMPOSITE appear as fine lines, which do not grow as they do in metals. Dark areas near the hole represent delamination.

"The most important thing is that these materials are naturally durable and more damage-resistant," Reifsnider notes. For instance, homogeneous materials such as metals lend themselves to cracking, but cracks simply won't grow in nonhomogeneous composites. Also, composites can change their properties greatly and still be safe, he says. Even if the stiffness drops about 40 percent, composites can still maintain their strength, unlike aluminum and steel, Reifsnider observes.

In the real world of aircraft design, engineers also rely on composites they feel comfortable handling. "We deal with a relatively limited set of fiber-reinforced composites," says Alan G. Miller, chief engineer of structures for Boeing Materials Technology. "We do use a limited set of chemistries, and we select chemistries that we think will be stable, based on experience over the years in various other industries." In addition to data from sources such as MRLife, Miller says aircraft manufacturers conduct accelerated testing (exposing the material to repeated cycles of heat, humidity and

AP PHOTO/CANARD AVIATORS

LONG-EZ is made from composites.

other stresses) and subject composites to unrealistically extreme conditions. "Are these things intrinsically flawed? The answer to that is clearly not," he insists.

So what about those high-profile accidents? The final word on the Denver and the Gleick crashes won't be in for a few months, but preliminary investigations do not implicate composites (some reports speculated that, in Denver's case, the plane ran out of fuel or hit a bird). But the U.S. Air Force says it knows what caused its accident: mechanics failed to reinstall four out of five fasteners on the wing after a routine maintenance inspection. No matter how tough composites are, it appears they are no match for human error. *—Phil Scott in New York City*

EVIDENCE DETECTION

SCENE OF THE CRIME

High-tech ways to see and collect evidence

Roughly 12 years ago forensic technician Garold L. Gresham was investigating a woman's murder. There were no solid leads for detectives, except that the killer's method suggested that a man had committed the act. As Gresham sifted through the crime scene—identifying fibers and fingerprints and typing blood—he discovered minute, red flecks of a hard, shiny material near the body. Under a microscope, they looked like paint, but with a difference. Gresham noticed that they were slightly ridged across their surfaces, much like the natural contours on fingernails. He was looking at nail polish, but it wasn't the victim's. He alerted the police that in searching for a male suspect, they were probably barking up the wrong tree.

Today if Gresham, now an advisory scientist at the Idaho National Engineering and Environmental Laboratory, could revisit that crime scene, he would most likely just load those flecks into a new tool that he and his colleague Gary S. Groenewold are developing with the University of Montana and the National Institute of Justice: a secondary ion mass spectrometer (SIMS) gun. The device would blast the surface of the chips with large ions that behave like atomic-size jackhammers, prying molecules from the surface and capturing them for characterization. Not only would Gresham have been able to tell that his sample was nail polish, he could also have identified nonvolatile organic chemicals, such as the soap with which the suspect washed her hands or perhaps residue of the perfume she wore.

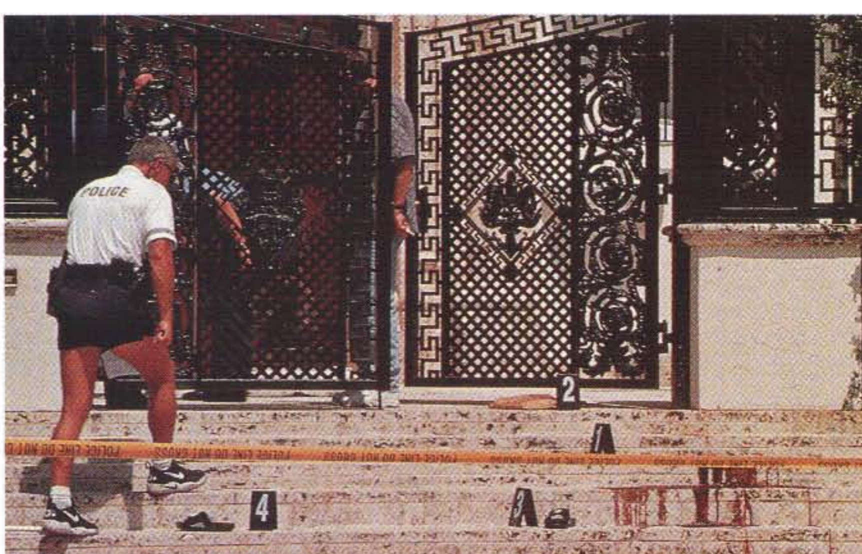
Groenewold found out—in his own lab—exactly how accurate the SIMS gun can be. "We kept reading an intense signature at mass 100. We thought it might be a six-carbon amine but couldn't figure out what the devil it was," he recalls. It turned out to be cyclohexylamine, an antiscaling agent used in boilers. "The cyclohexylamine was being distributed in the air at the concentration of about 200 parts per billion, but even such a small concentration was enough to create a chemical spike," he adds.

The SIMS gun is just one of several advances in evidence detection that may soon put a damper on the perfect crime. Researchers at Oak Ridge National Laboratory led by Michael E. Sigmond are working to improve explosive residue retrieval. Sigmond's team, with the National Bureau of Alcohol, Tobacco and Firearms (ATF), will soon test a dry sampling wipe that may allow blast experts to detect traces of explosives on surfaces. The dry wipe, made from a thermally stable material, could be loaded directly into a gas chromatograph and heated, allowing the entire gas phase of the sample to ascend directly into a column for analysis. Traditionally, a solvent is used to dissolve the residues off surfaces, which means that at least part of a sample is lost.

Mary Lou Fultz, chief of the ATF's Forensic Science Laboratory in Rockville, Md., is looking forward to ditching the solvents. "They're time-consuming. They cost a lot of money, and you have to make sure they're disposed of properly. The dry method will cut down on evidence handling" and hence limit sample contamination, she says.

Other advances are cutting down on evidence contamination and making collection more efficient by helping the police see what is ordinarily invisible—before it is accidentally stepped on or smudged. Consider this logistical nightmare: after the worst mass murder in Albuquerque's history, in which an armed robber executed three employees at a video store, eight detectives spent six painstaking hours dusting the entire video display on the store's south wall for fingerprints. "There could have been prints anywhere on that wall," says Detective Josephine "J. D." Herrera.

Enter Colin L. Smithpeter of Sandia National Laboratories, who is working on a project that would help detectives quickly spot evidence. His new tool will combine a device the police already use—an ultraviolet source called a blue-light special, which can pick up the natural fluorescence of body fluids in a darkened room—with heterodyning, the process by which two frequencies are combined and efficiently boosted to make them more easily detected (the



GREENWOOD *Liaison International*

EVIDENCE IDENTIFICATION, *such as that shown here after fashion designer Gianni Versace's murder, could be easier with new technology.*

GARY S. GROENEWOLD

CARPET FIBER analyzed by a SIMS gun glows because of organic molecules on its surface.

method that makes radio possible). A special camera implements the principle to separate the fluorescing wavelengths of body fluids from the background daylight. So Sandia's instrument would allow detectives to approach a scene in broad daylight and, with goggles equipped with shutters, watch fingerprints, semen, blood and other evidence blink back at them.

But the design needs more fine-tuning. "It's impossible to say at this point whether this thing is going to be able to differentiate old fluorescence from new," Smithpeter says. "It's just going to take some experimentation."

Unfortunately, none of these improvements in evidence detection and sampling guarantees an open-and-shut case. But they do promise to lessen evidence contamination and should shave time and money from the labors of overburdened precincts and courtrooms.

—*Brenda Dekoker Goodman in Albuquerque, N.M.*

MACHINING

FAKE IT BEFORE YOU MAKE IT

Virtual machine tools may transform manufacturing

Manufacturers of the machine tools that make the parts for a Ford Explorer or a Honda Accord still often rely on trial and error to correct misalignments in a grinding wheel or cutting machine. "Typically, you just try one thing, and if it doesn't work you try something else," says William W. Pflager, manager of research and development for Landis Gardner, a machine tool supplier to the automotive industry. It may be enough to make a small adjustment to the wheel that grinds the cylindrical shape of the main bearing of an automobile's crankshaft. But the imprecision of this method can often lead to costly delays when the machine is working a part with a more complicated geometry.

A collaboration among academia, industry and government has created simulation software that may be able to anticipate how well a tool forms a part even before the machine is built—or else it can diagnose faults in machines in operation. Machining variation analysis (MVA) can simulate the workings of machine tools that move with five degrees of freedom (three spatial dimensions as well as rotation about two axes). MVA creates a model of a virtual part, along with divergences from design specifications for machining, grinding, assembly, placement of parts, or other processes.

Previous simulation algorithms could only estimate machine errors at a single point on the part, whereas MVA creates a representation of anomalies across the entire surface. The simulation—a collaborative effort of researchers from the Massachusetts Institute of Technology, the National Institute of Standards and Technology and Landis Gardner—consists of a set of algorithms that can model any tool, from an engine valve grinder to the photolithographic machines used to pattern microchip circuits.

MVA works by incorporating three-dimensional data about the requirements for the desired geometry of the part. It uses this information to compute the "swept envelope": the precise path of the tool as it removes material from the workpiece. Other swept-envelope algorithms can provide only a rough approximation of the tool trajectory.

Another advance was to use the simulation to show how a dozen or so different types of machine errors—deviations from specification for a part's squareness or cylindrical form, for instance—can combine to affect the final shape of the part. Daniel D. Frey, assistant director of the System Design and Management program at M.I.T. and MVA's primary architect, says the flaws in a virtual part could be modeled only after he realized that the machine-operating characteristics that lead to errors could usually be modeled separately.

Machine misalignments that can affect the roundness of a part might be caused by unwanted motions in the spindle that holds the grinding wheel and by slight movements of the table on which a part rests, which result from heat generated by the grinding process. Both contributions to the roundness error can be characterized as independent variables, without having to worry about how one interacts with another.

After the variables are simulated, they need to be combined to create a representation of the shape of the final part. But these measurements do not necessarily add up in linear fashion, making calculation difficult. For instance, a screw may deviate from the specification for its desired roundness by 1.9 microns because of the spindle misalignment; at the same time it may be off by 2.0 microns because of deflections caused by the heat generated by the machine. But the two errors might add up to 3.5 microns, not the 3.9 microns that would result if the two variables summed linearly.

As a result, researchers rely on a probability technique, Monte Carlo analysis, which takes samples of random variables that represent the cause of an error. The random numbers can then be processed and summed to yield a composite "error signature" that represents a close approximation of the final shape of the part. These steps are repeated thousands of times. Each of the virtual parts produced during a given iteration is measured and incorporated into a statistical profile, yielding an overall indication of, say, roundness or squareness error. "These patterns of error are like the pathology of a disease," Frey says. "You can match them to a database of known errors and make a diagnosis of a problem."

The benefits of virtual machine tools have only begun to be deployed. Landis Gardner, which is based in Waynesboro, Pa., uses MVA to calculate error signatures for a machine that grinds the bearings that connect automotive pistons to crankshafts. (It has yet to employ routinely the sophisticated probability simulations of MVA.) Hughes Research Laboratories has experimented with MVA to estimate faulty placement of wire leads attached to a printed circuit board. In coming years, Frey anticipates varied applications, from reducing the need for parts inspections to finding the machine's "sweet spot," the best placement of a part in a machine tool. Virtual machine tools may thus yield real benefits. —*Gary Stix*

Downloading as a Crime

On December 16, 1997, President Bill Clinton signed into law the No Electronic Theft (NET) Act. This cheerful piece of legislation makes it a federal crime to distribute or possess unauthorized electronic copies of copyrighted material valued over $1,000, even when no profit is involved. The bill defines three levels of violations, depending on the value of the work and the number of past offenses. Possession of 10 or more illegal electronic copies worth more than $2,500 could land you six years in prison and a $250,000 fine. "You'd be better off going out and shooting somebody," quips David J. Farber, professor of telecommunications at the University of Pennsylvania. "The penalty is less."

Farber was also a leading signatory to a letter from the Association of Computing Machinery sent to President Clinton, asking him to veto the bill. The ACM's objections had to do with the potential for damaging free communication among scientists, because the bill does not contain traditional fair-use exemptions, such as those allowing photocopying by libraries and academic institutions or quotation for purposes of review or criticism, or first sale, which allows you to loan or sell a secondhand book.

Arguments over the correct balance between rewarding authors and publishers via copyright and the public's right of access to information are nothing new. In 1841, for example, in a House of Commons debate over the extension of copyright from 28 to 60 years, Thomas Babington Macaulay called copyright "a private tax on the innocent pleasure of reading" and "a tax on readers for the purpose of giving a bounty to writers" that should not be allowed to last a day longer than necessary for remunerating authors enough to keep them in business.

But since the arrival of the digital age, these arguments have taken a nasty turn, partly because it is so easy to copy and distribute digital material and partly because the technology is developing that would allow every use to be monitored and charged. For example, Mark J. Stefik of the Xerox Palo Alto Research Center is working on "trusted systems" that would make it possible to divide rights finely, so you could buy, say, reading rights for an article on the Internet but not downloading and printing rights. Your printer might be able to mark printouts undetectably, charging you for the job and sending out an electronic payment. Already digital watermarking—using a cryptographic technique called steganography to hide ownership and copyright information in a picture or text file—is being deployed to make it easier to identify infringers. All these technologies will make possible an unbundling of rights that up until now have been taken for granted. The habit of mind that finds this approach desirable has been called copyright maximalism by MacArthur award recipient Pamela Samuelson, professor of information management and of law at the University of California at Berkeley.

The NET Act does not contain traditional fair-use exemptions, such as those allowing photocopying by libraries or quotation for purposes of review or criticism.

In fact, the NET Act is only the first of several pieces of legislation before Congress, some seeking to place greater restrictions on the use and copying of digital information than exist in traditional media. One, the Digital Era Copyright Enhancement Act of 1997—introduced into the House on November 13, 1997, by Representatives Rick Boucher of Virginia and Tom Campbell of California—includes fair-use provisions; the other, introduced on July 29, 1997, on behalf of the Clinton administration by Representative Howard Coble of North Carolina and others, does not. In addition, the Coble et al. bill contains provisions making it illegal to provide or own technology that can defeat copyright-protection technology. The bills are intended as legislation that ratify the treaties passed by the diplomatic conference of the World Intellectual Property Organization in December 1996.

The NET Act had a different genesis; it was inspired by the dismissal of charges against Massachusetts Institute of Technology student David M. LaMacchia in 1994. He was charged with allowing the piracy of more than $1 million in business and entertainment software from an electronic bulletin board he ran on M.I.T.'s system. His attorneys successfully argued that LaMacchia did not profit from the site and that he himself did not upload, download or use the software available. The programs were supplied and retrieved by others over whom he had no control, and existing U.S. law did not cover this situation.

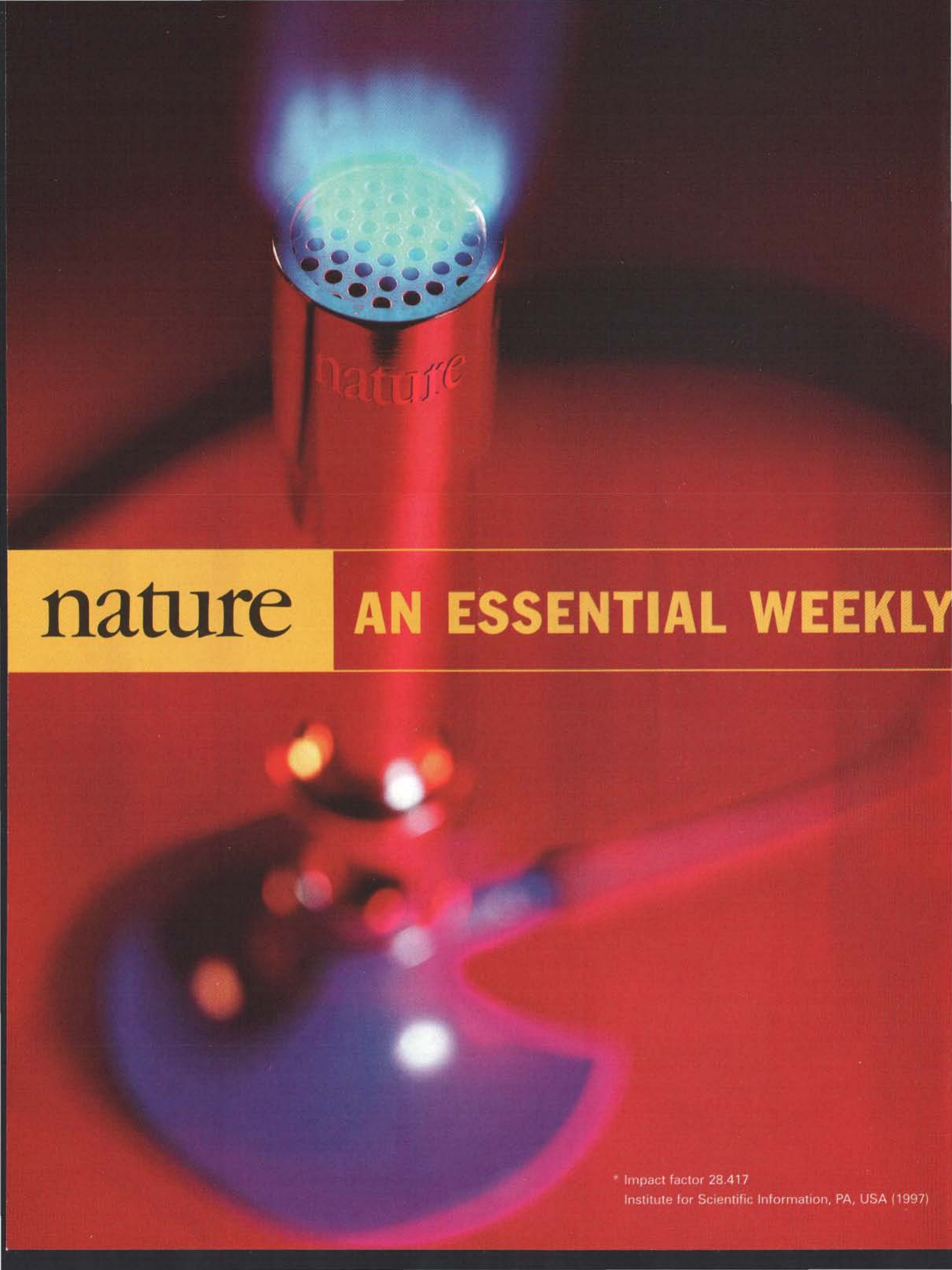
Now with the NET Act, it does. "It's an unfortunate piece of legislation," says Peter Jaszi of American University. Although the law will most likely not be misused in the way the ACM foresees, "the chilling effect that the risk of liability will generate is probably going to be significant and real," insists Jaszi, who is also co-founder of the Digital Future Coalition, a group of 39 public and private organizations that is backing the Boucher-Campbell bill and arguing for the extension of fair use into the digital environment.

In part, what it comes down to is that the Internet scares people, particularly copyright owners whose wealth is tied up in intellectual property. All over the Internet (and off it), large companies are attacking any use of their trademarked names they don't like, often apparently irrationally: Mattel is going after anyone who uses the name "Barbie"; movie studios threaten fan club sites that publish pictures, sounds and new fiction using their established characters; and, as *60 Minutes* reported in December 1997, McDonald's seems to think it is the sole owner of a name that is common to a large chunk of Scotland.

No one is arguing that the Internet doesn't pose a challenge to the traditional control that copyright owners had over their work. But even content providers themselves will be ill served by a regime under which they have to pay for each piece of information used in the production of new work. Ultimately, freedom of speech and of the press will mean nothing if everywhere information flows there are toll roads.

—Wendy M. Grossman in London

WENDY M. GROSSMAN is the author of net.wars, *published by New York University Press (1998).*



nature
nature
AN ESSENTIAL WEEKLY
* Impact factor 28.417
Institute for Scientific Information, PA, USA (1997)

The Bose-Einstein Condensate

Three years ago in a Colorado laboratory, scientists realized a long-standing dream, bringing the quantum world closer to the one of everyday experience

by Eric A. Cornell and Carl E. Wieman

In June 1995 our research group at the Joint Institute for Laboratory Astrophysics (now called JILA) in Boulder, Colo., succeeded in creating a minuscule but marvelous droplet. By cooling 2,000 rubidium atoms to a temperature less than 100 billionths of a degree above absolute zero (100 billionths of a degree kelvin), we caused the atoms to lose for a full 10 seconds their individual identities and behave as though they were a single "superatom." The atoms' physical properties, such as their motions, became identical to one another. This Bose-Einstein condensate (BEC), the first observed in a gas, can be thought of as the matter counterpart of the laser—except that in the condensate it is atoms, rather than photons, that dance in perfect unison.

Our short-lived, gelid sample was the experimental realization of a theoretical construct that has intrigued scientists ever since it was predicted some 73 years ago by the work of physicists Albert Einstein and Satyendra Nath Bose. At ordinary temperatures, the atoms of a gas are scattered throughout the container holding them. Some have high energies (high speeds); others have low ones. Expanding on Bose's work, Einstein showed that if a sample of atoms were cooled sufficiently, a large fraction of them would settle into the single lowest possible energy state in the container. In mathematical terms, their individual wave equations—which describe such physical characteristics of an atom as its position and velocity—would in effect merge, and each atom would become indistinguishable from any other.

Progress in creating Bose-Einstein condensates has sparked great interest in the physics community and has even generated coverage in the mainstream press. At first, some of the attention derived from the drama inherent in the decades-long quest to prove Einstein's theory. But most of the fascination now stems from the fact that the condensate offers a macroscopic window into the strange world of quantum mechanics, the theory of matter based on the observation that elementary particles, such as electrons, have wave properties. Quantum mechanics, which encompasses the famous Heisenberg uncertainty principle, uses these wavelike properties to describe the structure and interactions of matter.

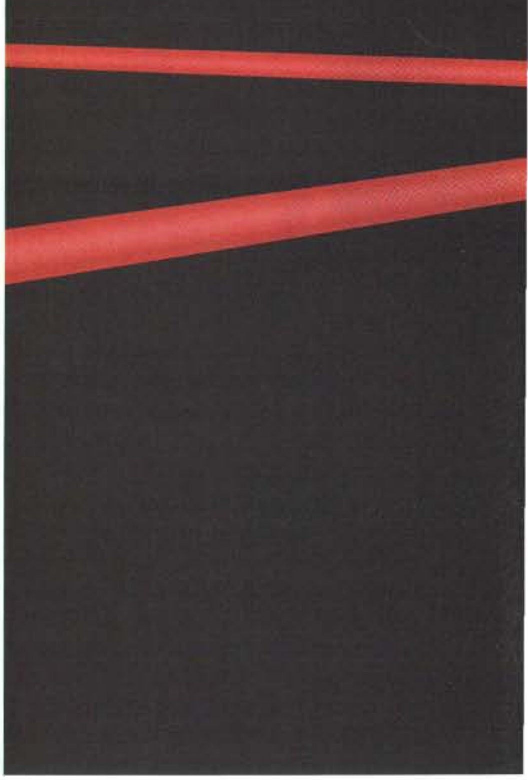
We can rarely observe the effects of quantum mechanics in the behavior of a macroscopic amount of material. In ordinary, so-called bulk matter, the incoherent contributions of the uncountably large number of constituent particles obscure the wave nature of quantum mechanics, and we can only infer its effects. But in Bose condensation, the wave nature of each atom is precisely in phase with that of every other. Quantum-mechanical waves extend across the sample of condensate and can be observed with the naked eye. The submicroscopic thus becomes macroscopic.

New Light on Old Paradoxes

The creation of Bose-Einstein condensates has cast new light on longstanding paradoxes of quantum mechanics. For example, if two or more atoms are in a single quantum-mechanical state, as they are in a condensate, it is fundamentally impossible to distinguish them by any measurement. The two atoms occupy the same volume of space, move at the identical speed, scatter light of the same color and so on.

Nothing in our experience, based as it is on familiarity with matter at normal temperatures, helps us comprehend this paradox. That is because at normal temperatures and at the size scales we are all familiar with, it is possible to de-

ATOMIC TRAP cools by means of two different mechanisms. First, six laser beams (*red*) cool atoms, initially at room temperature, while corralling them toward the center of an evacuated glass box. Next, the laser beams are turned off, and the magnetic coils (*copper*) are energized. Current flowing through the coils generates a magnetic field that further confines most of the atoms while allowing the energetic ones to escape. Thus, the average energy of the remaining atoms decreases, making the sample colder and even more closely confined to the center of the trap. Ultimately, many of the atoms attain the lowest possible energy state allowed by quantum mechanics and become a single entity known as a Bose-Einstein condensate.

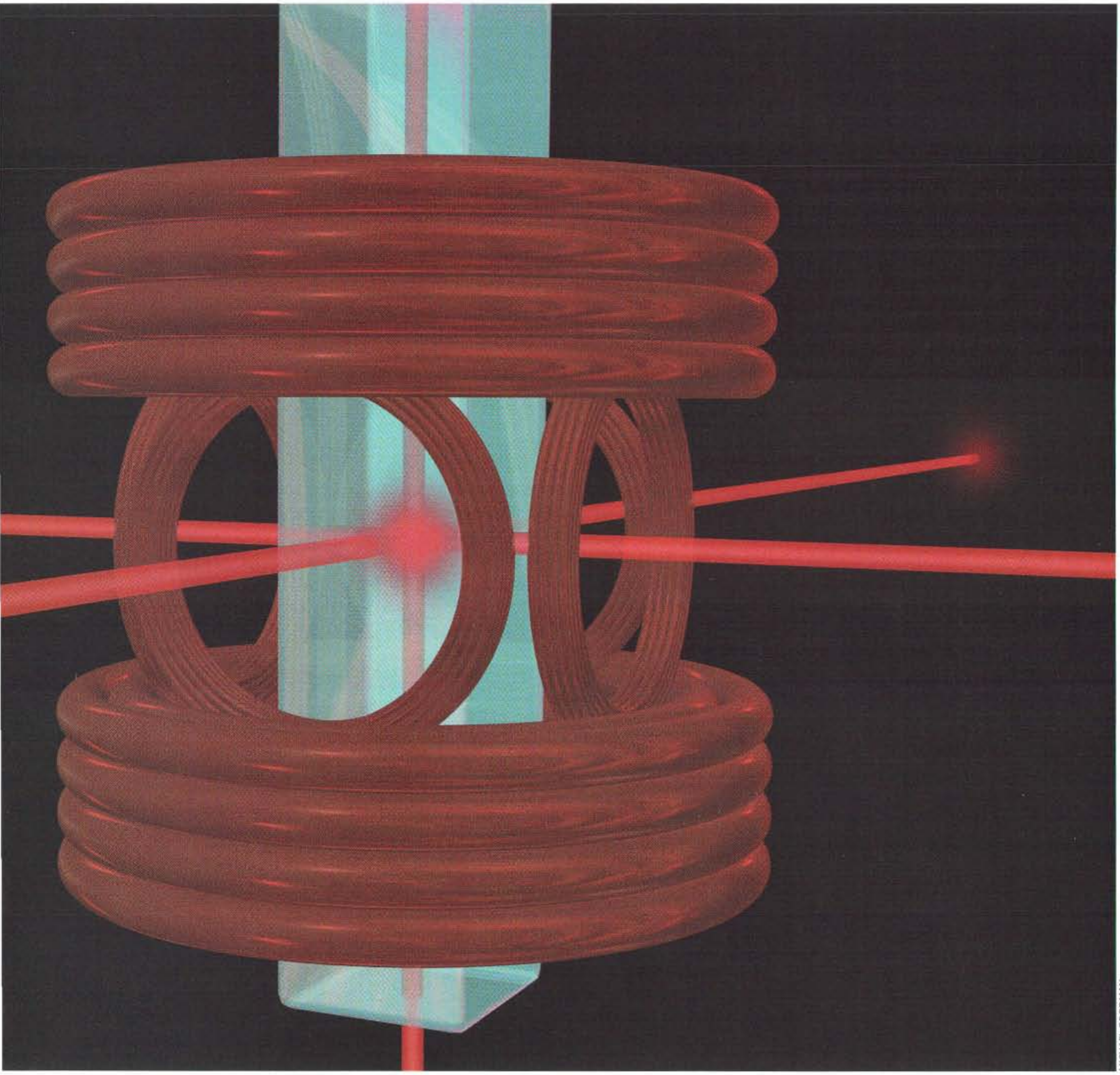


scribe the position and motion of each and every object in a collection of objects. The numbered Ping-Pong balls bouncing in a rotating drum used to select lottery numbers exemplify the motions describable by classical mechanics.

At extremely low temperatures or at small size scales, on the other hand, the usefulness of classical mechanics begins to wane. The crisp analogy of atoms as Ping-Pong balls begins to blur. We cannot know the exact position of each atom, which is better thought of as a blurry spot. This spot—known as a wave packet—is the region of space in which we can expect to find the atom. As a collection of atoms becomes colder, the size of each wave packet grows. As long as each wave packet is spatially separated from the others, it is possible, at least in principle, to tell atoms apart. When the temperature becomes sufficiently low, however, each atom's wave packet begins to overlap with those of neighboring atoms. When this happens, the atoms "Bose-condense" into the lowest possible energy state, and the wave packets coalesce into a single, macroscopic packet. The atoms undergo a quantum identity crisis: we can no longer distinguish one atom from another.

The current excitement over these condensates contrasts sharply with the reaction to Einstein's discovery in 1925 that they could exist. Perhaps because of the impossibility then of reaching the required temperatures—less than a millionth of a degree kelvin—the hypothesized gaseous condensate was considered a curiosity of questionable validity and little physical significance. For perspective, even the coldest depths of intergalactic space are millions of times too hot for Bose condensation.

In the intervening decades, however,



MICHAEL GOODMAN

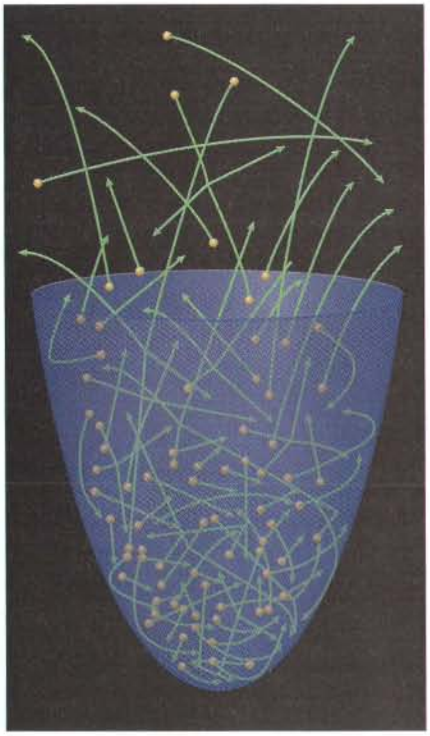
Bose condensation came back into fashion. Physicists realized that the concept could explain superfluidity in liquid helium, which occurs at much higher temperatures than gaseous Bose condensation. Below 2.2 kelvins, the viscosity of liquid helium completely disappears—putting the "super" in superfluidity.

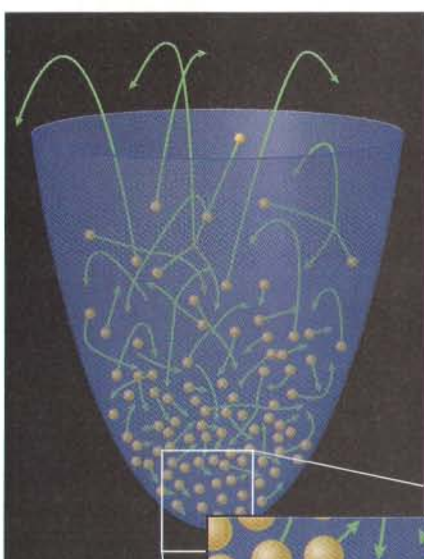
Not until the late 1970s did refrigeration technology advance to the point that physicists could entertain the notion of creating something like Einstein's original concept of a BEC in a gas. Laboratory workers at M.I.T., the University of Amsterdam, the University of British Columbia and Cornell University had to confront a fundamental difficulty. To achieve such a BEC, they had to cool the gas to far below the temperature at which the atoms would normally freeze into a solid. In other words, they had to create a supersaturated gas. Their expectation was that hydrogen would supersaturate, because the gas was known to resist the atom-by-atom clumping that precedes bulk freezing.

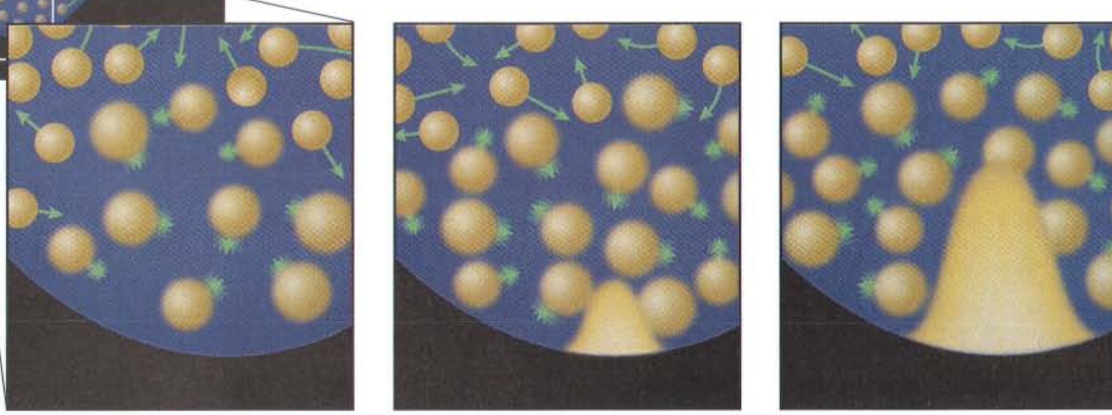
Although these investigators have not yet succeeded in creating a Bose-Einstein condensate with hydrogen, they did develop a much better understanding of the difficulties and found clever approaches for attacking them, which benefited us. In 1989, inspired by the hydrogen work and encouraged by our own research on the use of lasers to trap and cool alkali atoms, we began to suspect that these atoms, which include cesium, rubidium and sodium, would make much better candidates than hydrogen for producing a Bose condensate. Although the clumping properties of cesium, rubidium and sodium are not superior to those of hydrogen, the rate at which those atoms transform themselves into condensate is much faster than the rate for hydrogen atoms. These much larger atoms bounce off one another at much higher rates, sharing energy among themselves more quickly, which allows the condensate to form before clumping can occur.

Also, it looked as if it might be relatively easy and inexpensive to get these atoms very cold by combining ingenious techniques developed for laser cooling and trapping of alkali atoms with the techniques for magnetic trapping and evaporative cooling developed by the researchers working with hydrogen. These ideas were developed in a series of discussions with our friend and former teacher, Daniel Kleppner, the co-leader of a group at M.I.T. that is attempting to create a condensate with hydrogen.

Our hypothesis about alkali atoms was ultimately fruitful. Just a few months after we succeeded with rubidium, Wolfgang Ketterle's group at M.I.T. produced a Bose condensate with sodium atoms; since that time, Ketterle's team has succeeded in creating a condensate with 10 million atoms. At the time of this writing, there are at least seven teams producing condensates. Besides our own group, others working with rubidium are Daniel J. Heinzen of the University of Texas at Austin, Gerhard Rempe of the University of Konstanz in Germany and Mark Kasevich of Yale University. In sodium, besides Ketterle's at M.I.T., there is a group led by Lene Vestergaard Hau of the Rowland Institute for Science in Cambridge,

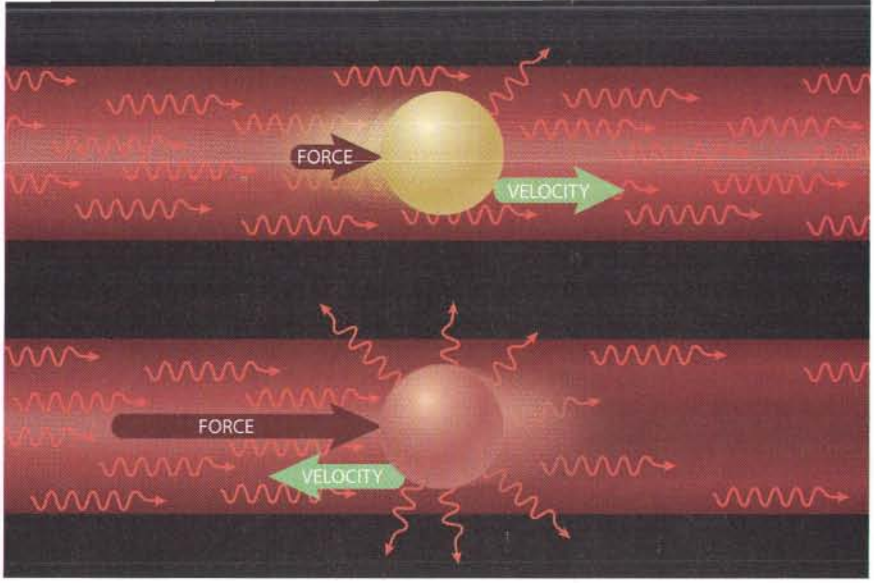






EVAPORATIVE COOLING occurs in a magnetic trap, which can be thought of as a deep bowl (*blue*). The most energetic atoms, depicted with the longest green trajectory arrows, escape from the bowl (*above, left*). Those that remain collide with one another frequently, apportioning out the remaining energy (*left*). Eventually, the atoms move so slowly and are so closely packed at the bottom of the bowl that their quantum nature becomes more pronounced. So-called wave packets, representing the region where each atom is likely to be found, become less distinct and begin to overlap (*below, left*). Ultimately, two atoms collide, and one is left as close to stationary as is allowed by Heisenberg's uncertainty principle. This event triggers an avalanche of atoms piling up in the lowest energy state of the trap, merging into the single ground-state blob that is a Bose-Einstein condensate (*below, center* and *right*).

MICHAEL GOODMAN



LASER COOLING of an atom makes use of the pressure, or force, exerted by repeated photon impacts. An atom moving against a laser beam encounters a higher frequency than an atom moving with the same beam. In cooling, the frequency of the beam is adjusted so that an atom moving into the beam scatters many more photons than an atom moving away from the beam. The net effect is to reduce the speed and thus cool the atom.

Mass. At Rice University Randall G. Hulet has succeeded in creating a condensate with lithium.

All these teams are using the same basic apparatus. As with any kind of refrigeration, the chilling of atoms requires a method of removing heat and also of insulating the chilled sample from its surroundings. Both functions are accomplished in each of two steps. In the first, the force of laser light on the atoms both cools and insulates them. In the second, we use magnetic fields to insulate, and we cool by evaporation.

Laser Cooling and Trapping

The heart of our apparatus is a small glass box with some coils of wire around it [*see illustration on pages 26 and 27*]. We completely evacuate the cell, producing in effect a superefficient thermos bottle. Next, we let in a tiny amount of rubidium gas. Six beams of laser light intersect in the middle of the box, converging on the gas. The laser light need not be intense, so we obtain it from inexpensive diode lasers, similar to those found in compact-disc players.

We adjust the frequency of the laser radiation so that the atoms absorb it and then reradiate photons. An atom can absorb and reradiate many millions of photons each second, and with each one, the atom receives a minuscule kick in the direction the absorbed photon is moving. These kicks are called radiation pressure. The trick to laser cooling is to get the atom to absorb mainly photons that are traveling in the direction opposite that of the atom's motion, thereby slowing the atom down (cooling it, in other words). We accomplish this feat by carefully adjusting the frequency of the laser light relative to the frequency of the light absorbed by the atoms [*see illustration above*].

In this setup, we use laser light not only to cool the atoms but also to "trap" them, keeping them away from the room-temperature walls of the cell. In fact, the two laser applications are similar. With trapping, we use the radiation pressure to oppose the tendency of the atoms to drift away from the center of the cell. A weak magnetic field tunes the resonance of the atom to absorb preferentially from the laser beam that is pointing toward the center of the cell (recall that six laser beams intersect at the center of the cell). The net effect is that all the atoms are pushed toward one spot and are held there just by the force of the laser light.

These techniques fill our laser trap in one minute with 10 million atoms captured from the room-temperature rubidium vapor in the cell. These trapped atoms are at a temperature of about 40 millionths of a degree above absolute zero—an extraordinarily low temperature by most standards but still 100 times too hot to form a BEC. In the presence of the laser light, the unavoidable random jostling the atoms receive from the impact of individual light photons keeps the atoms from getting any colder or denser.

To get around the limitations imposed by those random photon impacts, we turn off the lasers at this point and activate the second stage of the cooling process. This stage is based on the magnetic-trapping and evaporative-cooling technology developed in the quest to achieve a condensate with hydrogen atoms. A magnetic trap exploits the fact that each atom acts like a tiny bar magnet and thus is subjected to a force when placed in a magnetic field [*see illustration on opposite page*]. By carefully controlling the shape of the magnetic field and making it relatively strong, we can use the field to hold the atoms, which move around inside the field much like balls rolling about inside a deep bowl. In evaporative cooling, the most energetic atoms escape from this magnetic bowl. When they do, they carry away more than their share of the energy, leaving the remaining atoms colder.

The analogy here is to cooling coffee. The most energetic water molecules leap out of the cup into the room (as steam), thereby reducing the average energy of the liquid that is left in the cup. Meanwhile countless collisions among the remaining molecules in the cup apportion out the remaining energy among all those molecules. Our cloud of magnetically trapped atoms is at a much lower density than water molecules in a cup. So the primary experimental challenge we faced for five years was how to get the atoms to collide with one another enough times to share the energy before they were knocked out of the trap by a collision with one of the untrapped, room-temperature atoms remaining in our glass cell.

Many small improvements, rather than a single breakthrough, solved this problem. For instance, before assembling the cell and its connected vacuum pump, we took extreme care in cleaning each part, because any remaining residues from our hands on an inside surface would emit vapors that would degrade the vacuum. Also, we made sure that the tiny amount of rubidium vapor remaining in the cell was as small as it could be while providing a sufficient number of atoms to fill the optical trap.

Incremental steps such as these helped but still left us well shy of the density needed to get the evaporative cooling under way. The basic problem was the effectiveness of the magnetic trap. Although the magnetic fields that make

up the confining magnetic "bowl" can be quite strong, the little "bar magnet" inside each individual atom is weak. This characteristic makes it difficult to push the atom around with a magnetic field, even if the atom is moving quite slowly (as are our laser-cooled atoms).

In 1994 we finally confronted the need to build a magnetic trap with a narrower, deeper bowl. Our quickly built, narrow-and-deep magnetic trap proved to be the final piece needed to cool evaporatively the rubidium atoms into a condensate. As it turns out, our particular trap design was hardly a unique solution. Currently there are almost as many different magnetic trap configurations as there are groups studying these condensates.

Shadow Snapshot of a "Superatom"

How do we know that we have in fact produced a Bose-Einstein condensate? To observe the cloud of cooled atoms, we take a so-called shadow snapshot with a flash of laser light. Because the atoms sink to the bottom of the magnetic bowl as they cool, the cold cloud is too small to see easily. To make it larger, we turn off the confining magnetic fields, allowing the atoms to fly out freely in all directions. After about 0.1 second, we illuminate the now expanded cloud with a flash of laser light. The atoms scatter this light out of the beam, casting a shadow that we observe with a video camera. From this shadow, we can determine the distribution of velocities of the atoms in the original trapped cloud. The velocity measurement also gives us the temperature of the sample.

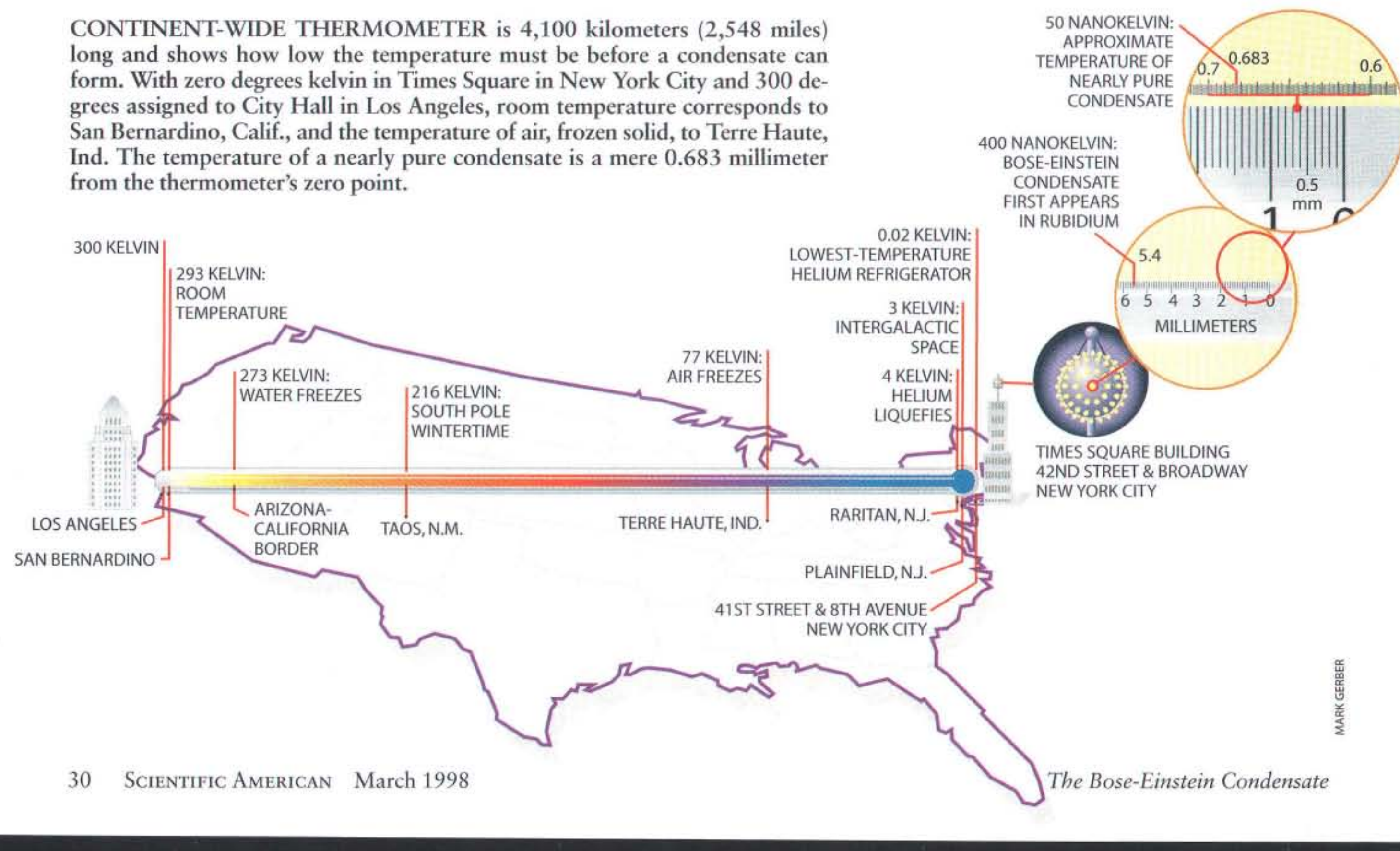
In the plot of the velocity distribution [*see illustration on opposite page*], the condensate appears as a dorsal-fin-shaped peak. The condensate atoms have the smallest possible velocity and thus remain in a dense cluster in the center of the cloud after it has expanded. This photograph of a condensate is further proof that there is something wrong with classical mechanics. The condensate forms with the lowest possible energy. In classical mechanics, "lowest energy" means that the atoms should be at the center of the trap and motionless, which would appear as an infinitely narrow and tall peak in our image. The peak differs from this classical conception because of quantum effects that can be summed up in three words: Heisenberg's uncertainty principle.

The uncertainty principle puts limits on what is knowable about anything, including atoms. The more precisely you know an atom's location, the less well you can know its velocity, and vice versa. That is why the condensate peak is not infinitely narrow. If it were, we would know that the atoms were in the exact center of the trap and had exactly zero energy. According to the uncertainty principle, we cannot know both these things simultaneously.

Einstein's theory requires that the atoms in a condensate have energy that is as low as possible, whereas Heisenberg's uncertainty principle forbids them from being at the very bottom of the trap. Quantum mechanics resolves this conflict by postulating that the energy of an atom in any container, including our trap, can only be one of a set of discrete, allowed values—and the lowest of these values is not quite zero. This lowest allowed energy is called the zero-point energy, because even atoms whose temperature is exactly zero have this minimum energy. Atoms with this energy move around slowly near—but not quite at—the center of the trap. The uncertainty principle and the other laws of quantum mechanics are normally seen only in the behavior of submicroscopic objects such as a single atom or smaller. The Bose-Einstein condensate therefore is a rare example of the uncertainty principle in action in the macroscopic world.

Bose-Einstein condensation of atoms is too new, and too different, for us to say if its usefulness will eventually extend beyond lecture demonstrations for quantum mechanics. Any discussion of practical applications for condensates must necessarily be speculative. Never-

CONTINENT-WIDE THERMOMETER is 4,100 kilometers (2,548 miles) long and shows how low the temperature must be before a condensate can form. With zero degrees kelvin in Times Square in New York City and 300 degrees assigned to City Hall in Los Angeles, room temperature corresponds to San Bernardino, Calif., and the temperature of air, frozen solid, to Terre Haute, Ind. The temperature of a nearly pure condensate is a mere 0.683 millimeter from the thermometer's zero point.



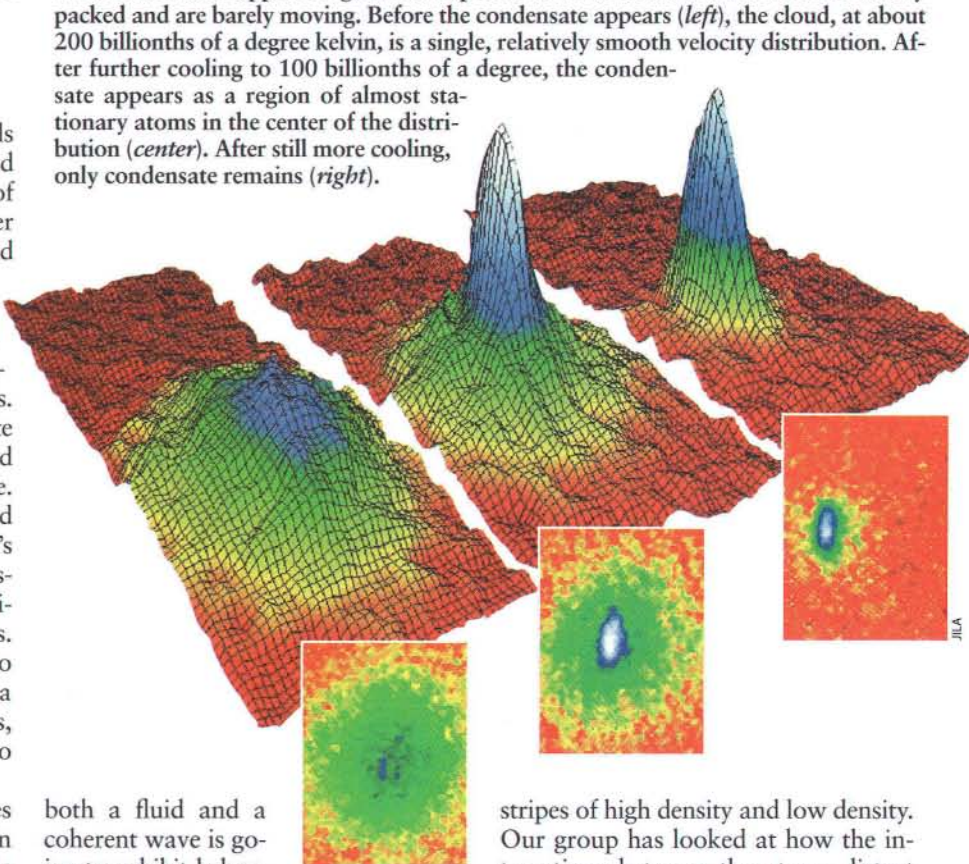
MARK GERBER

theless, our musings can be guided by a striking physical analogy: the atoms that make up a Bose condensate are in many ways the analogue to the photons that make up a laser beam.

The Ultimate in Precise Control?

Every photon in a laser beam travels in exactly the same direction and has the same frequency and phase of oscillation. This property makes laser light very easy to control precisely and leads to its utility in compact-disc players, laser printers and other appliances. Similarly, Bose condensation represents the ultimate in precise control—but for atoms rather than photons. The matter waves of a Bose condensate can be reflected, focused, diffracted and modulated in frequency and amplitude. This kind of control will very likely lead to improved timekeeping; the world's best clocks are already based on the oscillations of laser-cooled atoms. Applications may also turn up in other areas. In a flight of fancy, it is possible to imagine a beam of atoms focused to a spot only a millionth of a meter across, "airbrushing" a transistor directly onto an integrated circuit.

But for now, many of the properties of the Bose-Einstein condensate remain unknown. Of particular interest is the condensate's viscosity. The speculation now is that the viscosity will be vanishingly small, making the condensate a kind of "supergas," in which ripples and swirls, once excited, will never damp down. Another area of curiosity centers on a basic difference between laser light and a condensate. Laser beams are noninteracting—they can cross without affecting one another at all. A condensate, on the other hand, has some resistance to compression and some springiness—it is, in short, a fluid. A material that is both a fluid and a coherent wave is going to exhibit behavior that is rich, which is a physicist's way of saying that it is going to take a long time to figure out.



SHADOW IMAGE of a forming Bose-Einstein condensate was processed by a computer to show more clearly the distribution of velocities of atoms in the cold cloud. Top and bottom images show the same data but from different angles. In the upper set, where the surface appears highest corresponds to where the atoms are the most closely packed and are barely moving. Before the condensate appears (*left*), the cloud, at about 200 billionths of a degree kelvin, is a single, relatively smooth velocity distribution. After further cooling to 100 billionths of a degree, the condensate appears as a region of almost stationary atoms in the center of the distribution (*center*). After still more cooling, only condensate remains (*right*).

JILA

Meanwhile many groups have begun a variety of measurements on the condensates. In a lovely experiment, Ketterle's group has already shown that when two separate clouds of Bose condensate overlap, the result is a fringe pattern of alternating constructive and destructive interference, just as occurs with intersecting laser radiation. In the atom cloud, these regions appear respectively as stripes of high density and low density. Our group has looked at how the interactions between the atoms distort the shape of the atom cloud and the manner in which it quivers after we have "poked" it gently with magnetic fields. A number of other teams are now devising their own experiments to join in this work.

As the results begin to accrue from these and other experiments over the next several years, we will improve our understanding of this singular state of matter. As we do, the strange, fascinating quantum-mechanical world will come a little bit closer to our own. SA

The Authors

ERIC A. CORNELL and CARL E. WIEMAN are both fellows of JILA, the former Joint Institute for Laboratory Astrophysics, which is staffed by the National Institute of Standards and Technology (NIST) and the University of Colorado. Cornell, a physicist at NIST and a professor adjoint at the university, was co-leader, with Wieman, of the team at JILA that produced the first Bose-Einstein condensate in a gas. Wieman, a professor of physics at the university, is also known for his studies of the breakdown of symmetry in the interactions of elementary particles. The authors would like to thank their colleagues Michael Anderson, Michael Matthews and Jason Ensher for their work on the condensate project.

Further Reading

NEW MECHANISMS FOR LASER COOLING. William D. Phillips and Claude Cohen-Tannoudji in *Physics Today,* Vol. 43, pages 33–40; October 1990.

OBSERVATION OF BOSE-EINSTEIN CONDENSATION IN A DILUTE ATOMIC VAPOR. M. H. Anderson, J. R. Ensher, M. R. Matthews, C. E. Wieman and E. A. Cornell in *Science,* Vol. 269, pages 198–201; July 14, 1995.

BOSE-EINSTEIN CONDENSATION. Edited by A. Griffin, D. W. Snoke and S. Stringari. Cambridge University Press, 1995.

OBSERVATION OF INTERFERENCE BETWEEN TWO BOSE CONDENSATES. M. R. Andrews, C. G. Townsend, H.-J. Miesner, D. S. Durfee, D. M. Kurn and W. Ketterle in *Science,* Vol. 275, pages 637–641; January 31, 1997.

BOSE-EINSTEIN CONDENSATION. World Wide Web site available at http://www.colorado.edu/physics/2000/bec

The Challenge of Antibiotic Resistance

Certain bacterial infections now defy all antibiotics. The resistance problem may be reversible, but only if society begins to consider how the drugs affect "good" bacteria as well as "bad"

by Stuart B. Levy

Last year an event doctors had been fearing finally occurred. In three geographically separate patients, an often deadly bacterium, *Staphylococcus aureus*, responded poorly to a once reliable antidote—the antibiotic vancomycin. Fortunately, in those patients, the staph microbe remained susceptible to other drugs and was eradicated. But the appearance of *S. aureus* not readily cleared by vancomycin foreshadows trouble.

Worldwide, many strains of *S. aureus* are already resistant to all antibiotics except vancomycin. Emergence of forms lacking sensitivity to vancomycin signifies that variants untreatable by every known antibiotic are on their way. *S. aureus*, a major cause of hospital-acquired infections, has thus moved one step closer to becoming an unstoppable killer.

The looming threat of incurable *S. aureus* is just the latest twist in an international public health nightmare: increasing bacterial resistance to many antibiotics that once cured bacterial diseases readily. Ever since antibiotics became widely available in the 1940s, they have been hailed as miracle drugs—magic bullets able to eliminate bacteria without doing much harm to the cells of treated individuals. Yet with each passing decade, bacteria that defy not only single but multiple antibiotics—and therefore are extremely difficult to control—have become increasingly common.

What is more, strains of at least three bacterial species capable of causing life-threatening illnesses (*Enterococcus faecalis, Mycobacterium tuberculosis* and *Pseudomonas aeruginosa*) already evade every antibiotic in the clinician's armamentarium, a stockpile of more than 100 drugs. In part because of the rise in resistance to antibiotics, the death rates for some communicable diseases (such as tuberculosis) have started to rise again, after having declined in the industrial nations.

How did we end up in this worrisome, and worsening, situation? Several interacting processes are at fault. Analyses of them point to a number of actions that

Staphylococcus aureus

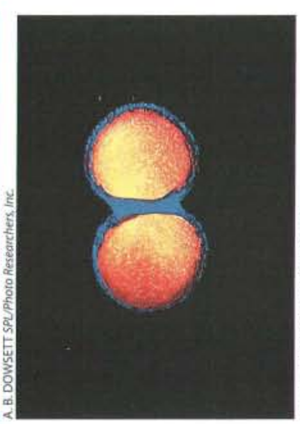
Causes blood poisoning, wound infections and pneumonia; in some hospitals, more than 60 percent of strains are resistant to methicillin; some are poised for resistance to all antibiotics (H/C; 1950s)

Acinetobacter

Causes blood poisoning in patients with compromised immunity (H, 1990s)

Enterococcus faecalis

Causes blood poisoning and urinary tract and wound infections in patients with compromised immunity; some multidrug-resistant strains are untreatable (H, 1980s)



Neisseria gonorrhoeae

Causes gonorrhea; multidrug resistance now limits therapy chiefly to cephalosporins (C; 1970s)

Haemophilus influenzae

Causes pneumonia, ear infections and meningitis, especially in children. Now largely preventable by vaccines (C; 1970s)

could help reverse the trend, if individuals, businesses and governments around the world can find the will to implement them.

One component of the solution is recognizing that bacteria are a natural, and needed, part of life. Bacteria, which are microscopic, single-cell entities, abound on inanimate surfaces and on parts of the body that make contact with the outer world, including the skin, the mucous membranes and the lining of the intestinal tract. Most live blamelessly. In fact, they often protect us from disease, because they compete with, and thus limit the proliferation of, pathogenic bacteria—the minority of species that can multiply aggressively (into the millions) and damage tissues or otherwise cause illness. The benign competitors can be important allies in the fight against antibiotic-resistant pathogens.

People should also realize that although antibiotics are needed to control bacterial infections, they can have broad, undesirable effects on microbial ecology. That is, they can produce long-lasting change in the kinds and proportions of bacteria—and the mix of antibiotic-resistant and antibiotic-susceptible types—not only in the treated individual but also in the environment and society at large. The compounds should thus be used only when they are truly needed, and they should not be administered for viral infections, over which they have no power.

A Bad Combination

Although many factors can influence whether bacteria in a person or in a community will become insensitive to an antibiotic, the two main forces are the prevalence of resistance genes (which give rise to proteins that shield bacteria from an antibiotic's effects) and the extent of antibiotic use. If the collective bacterial flora in a community have no genes conferring resistance to a given antibiotic, the antibiotic will successfully eliminate infection caused by any of the bacterial species in the collection. On the other hand, if the flora possess resistance genes and the community uses the drug persistently, bacteria able to defy eradication by the compound will emerge and multiply.

Antibiotic-resistant pathogens are not more virulent than susceptible ones: the same numbers of resistant and susceptible bacterial cells are required to produce disease. But the resistant forms are harder to destroy. Those that are slightly insensitive to an antibiotic can often be eliminated by using more of the drug; those that are highly resistant require other therapies.

To understand how resistance genes enable bacteria to survive an attack by an antibiotic, it helps to know exactly what antibiotics are and how they harm bacteria. Strictly speaking, the compounds are defined as natural substances (made by living organisms) that inhibit the growth, or proliferation, of bacteria or kill them directly. In practice, though, most commercial antibiotics have been chemically altered in the laboratory to improve their potency or to increase the range of species they affect. Here I will also use the term to encompass completely synthetic medicines, such as quinolones and sulfonamides, which technically fit under the broader rubric of antimicrobials.

Whatever their monikers, antibiotics, by inhibiting bacterial growth, give a host's immune defenses a chance to outflank the bugs that remain. The drugs typically retard bacterial proliferation by entering the microbes and interfering with the production of components needed to form new bacterial cells. For instance, the antibiotic tetracycline binds to ribosomes (internal structures that make new proteins) and, in so doing, impairs protein manufacture; penicillin and vancomycin impede proper synthesis of the bacterial cell wall.

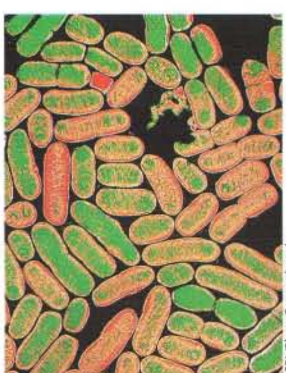
Certain resistance genes ward off destruction by giving rise to enzymes that

ROGUE'S GALLERY OF BACTERIA features some types having variants resistant to multiple antibiotics; multidrug-resistant bacteria are difficult and expensive to treat. Certain strains of the species described in red no longer respond to any antibiotics and produce incurable infections. Some of the bacteria cause infections mainly in hospitals (H) or mainly in the community (C); others, in both settings. The decade listed with each entry indicates the period when resistance first became a significant problem for patient care. The bacteria, which are microscopic, are highly magnified in these false-color images.

CAMR/A. B. DOWSETT SPL/Photo Researchers, Inc.

Mycobacterium tuberculosis

Causes tuberculosis; some multidrug-resistant strains are untreatable (H/C; 1970s)



E. GRAY SPL/Photo Researchers, Inc.

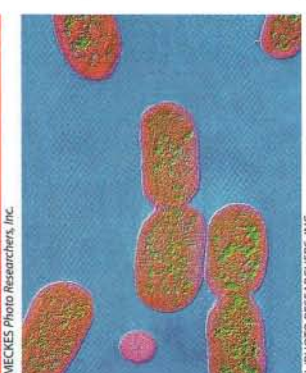
Escherichia coli

Causes urinary tract infections, blood poisoning, diarrhea and kidney failure; some strains that cause urinary tract infections are multidrug-resistant (H/C; 1960s)

OLIVER MECKES Photo Researchers, Inc.

Pseudomonas aeruginosa

Causes blood poisoning and pneumonia, especially in people with cystic fibrosis or compromised immunity; some multidrug-resistant strains are untreatable (H/C; 1960s)



CNRI/SPL/PHOTO RESEARCHERS, INC.

Shigella dysenteria

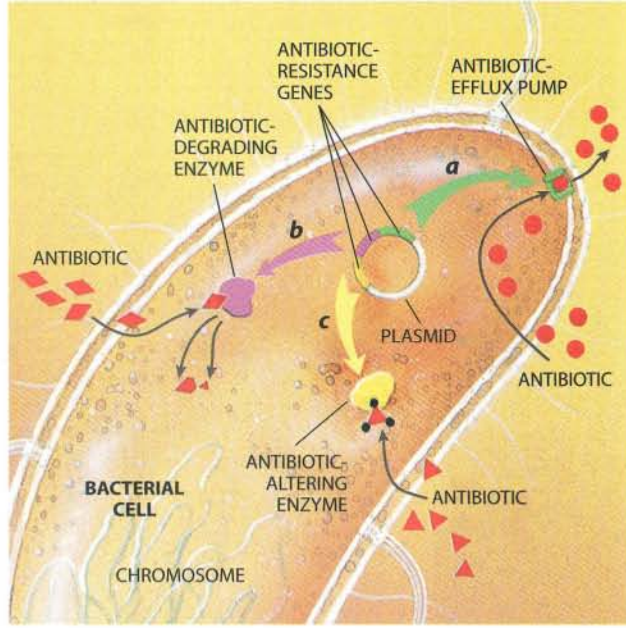
Causes dysentery (bloody diarrhea); resistant strains have led to epidemics, and some can be treated only by expensive fluoroquinolones, which are often unavailable in developing nations (C; 1960s)

CNRI/SPL/PHOTO RESEARCHERS, INC.

Streptococcus pneumoniae

Causes blood poisoning, middle ear infections, pneumonia and meningitis (C; 1970s)

ANTIBIOTIC-RESISTANT BACTERIA owe their drug insensitivity to resistance genes. For example, such genes might code for "efflux" pumps that eject antibiotics from cells (*a*). Or the genes might give rise to enzymes that degrade the antibiotics (*b*) or that chemically alter—and inactivate—the drugs (*c*). Resistance genes can reside on the bacterial chromosome or, more typically, on small rings of DNA called plasmids. Some of the genes are inherited, some emerge through random mutations in bacterial DNA, and some are imported from other bacteria.



degrade antibiotics or that chemically modify, and so inactivate, the drugs. Alternatively, some resistance genes cause bacteria to alter or replace molecules that are normally bound by an antibiotic—changes that essentially eliminate the drug's targets in bacterial cells. Bacteria might also eliminate entry ports for the drugs or, more effectively, may manufacture pumps that export antibiotics before the medicines have a chance to find their intracellular targets.

My Resistance Is Your Resistance

Bacteria can acquire resistance genes through a few routes. Many inherit the genes from their forerunners. Other times, genetic mutations, which occur readily in bacteria, will spontaneously produce a new resistance trait or will strengthen an existing one. And frequently, bacteria will gain a defense against an antibiotic by taking up resistance genes from other bacterial cells in the vicinity. Indeed, the exchange of genes is so pervasive that the entire bacterial world can be thought of as one huge multicellular organism in which the cells interchange their genes with ease.

Bacteria have evolved several ways to share their resistance traits with one another [see "Bacterial Gene Swapping in Nature," by Robert V. Miller; Scientific American, January]. Resistance genes commonly are carried on plasmids, tiny loops of DNA that can help bacteria survive various hazards in the environment. But the genes may also occur on the bacterial chromosome, the larger DNA molecule that stores the genes needed for the reproduction and routine maintenance of a bacterial cell.

DAN WAGNER (photograph); LAURIE GRACE (digital manipulation)

The Antibacterial Fad: A New Threat

Antibiotics are not the only antimicrobial substances being overexploited today. Use of antibacterial agents—compounds that kill or inhibit bacteria but are too toxic to be taken internally—has been skyrocketing as well. These compounds, also known as disinfectants and antiseptics, are applied to inanimate objects or to the skin.

Historically, most antibacterials were used in hospitals, where they were incorporated into soaps and surgical clothes to limit the spread of infections. More recently, however, those substances (including triclocarbon, triclosan and such quaternary ammonium compounds as benzalkonium chloride) have been mixed into soaps, lotions and dishwashing detergents meant for general consumers. They have also been impregnated into such items as toys, high chairs, mattress pads and cutting boards.

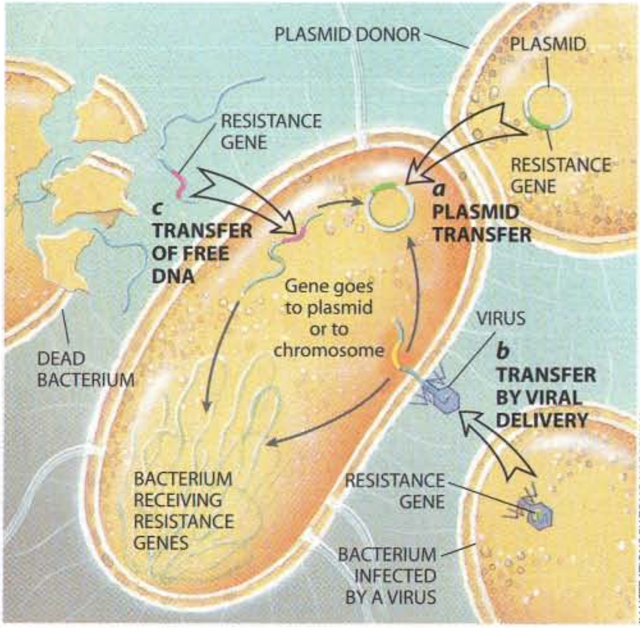
There is no evidence that the addition of antibacterials to such household products wards off infection. What is clear, however, is that the proliferation of products containing them raises public health concerns.

Like antibiotics, antibacterials can alter the mix of bacteria: they simultaneously kill susceptible bacteria and promote the growth of resistant strains. These resistant microbes may include bacteria that were present from the start. But they can also include ones that were unable to gain a foothold previously and are now able to thrive thanks to the destruction of competing microbes. I worry particularly about that second group—the interlopers—because once they have a chance to proliferate, some may become new agents of disease.

The potential overuse of antibacterials in the home is troubling on other grounds as well. Bacterial genes that confer resistance to antibacterials are sometimes carried on plasmids (circles of DNA) that also bear antibiotic-resistance genes. Hence, by promoting the growth of bacteria bearing such plasmids, antibacterials may actually foster double resistance—to antibiotics as well as antibacterials.

Routine housecleaning is surely necessary. But standard soaps and detergents (without added antibacterials) decrease the numbers of potentially troublesome bacteria perfectly well. Similarly, quickly evaporating chemicals—such as the old standbys of chlorine bleach, alcohol, ammonia and hydrogen peroxide—can be applied beneficially. They remove potentially disease-causing bacteria from, say, thermometers or utensils used to prepare raw meat for cooking, but they do not leave long-lasting residues that will continue to kill benign bacteria and increase the growth of resistant strains long after target pathogens have been removed.

If we go overboard and try to establish a sterile environment, we will find ourselves cohabiting with bacteria that are highly resistant to antibacterials and, possibly, to antibiotics. Then, when we really need to disinfect our homes and hands—as when a family member comes home from a hospital and is still vulnerable to infection—we will encounter mainly resistant bacteria. It is not inconceivable that with our excessive use of antibacterials and antibiotics, we will make our homes, like our hospitals, havens of ineradicable disease-producing bacteria. —*S.B.L.*



BACTERIA PICK UP RESISTANCE GENES from other bacterial cells in three main ways. Often they receive whole plasmids bearing one or more such genes from a donor cell (*a*). Other times, a virus will pick up a resistance gene from one bacterium and inject it into a different bacterial cell (*b*). Alternatively, bacteria sometimes scavenge gene-bearing snippets of DNA from dead cells in their vicinity (*c*). Genes obtained through viruses or from dead cells persist in their new owner if they become incorporated stably into the recipient's chromosome or into a plasmid.

Often one bacterium will pass resistance traits to others by giving them a useful plasmid. Resistance genes can also be transferred by viruses that occasionally extract a gene from one bacterial cell and inject it into a different one. In addition, after a bacterium dies and releases its contents into the environment, another will occasionally take up a liberated gene for itself.

In the last two situations, the gene will survive and provide protection from an antibiotic only if integrated stably into a plasmid or chromosome. Such integration occurs frequently, though, because resistance genes are often embedded in small units of DNA, called transposons, that readily hop into other DNA molecules. In a regrettable twist of fate for human beings, many bacteria play host to specialized transposons, termed integrons, that are like flypaper in their propensity for capturing new genes. These integrons can consist of several different resistance genes, which are passed to other bacteria as whole regiments of antibiotic-defying guerrillas.

Many bacteria possessed resistance genes even before commercial antibiotics came into use. Scientists do not know exactly why these genes evolved and were maintained. A logical argument holds that natural antibiotics were initially elaborated as the result of chance genetic mutations. Then the compounds, which turned out to eliminate competitors, enabled the manufacturers to survive and proliferate—if they were also lucky enough to possess genes that protected them from their own chemical weapons. Later, these protective genes found their way into other species, some of which were pathogenic.

Regardless of how bacteria acquire resistance genes today, commercial antibiotics can select for—promote the survival and propagation of—antibiotic-resistant strains. In other words, by encouraging the growth of resistant pathogens, an antibiotic can actually contribute to its own undoing.

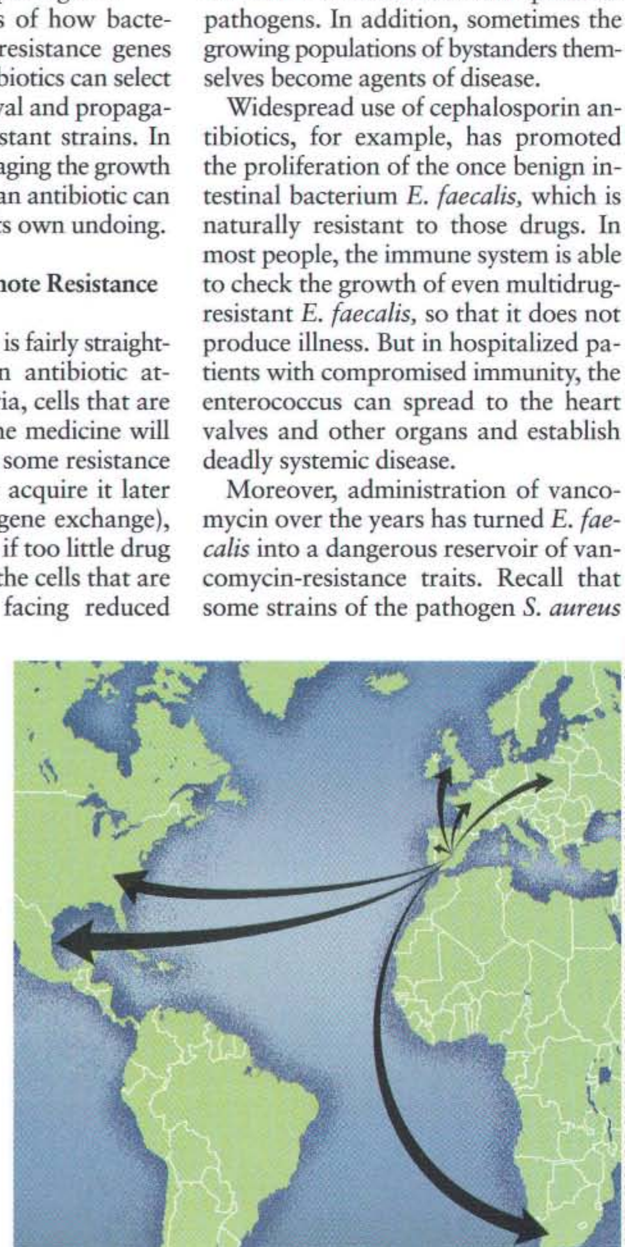
How Antibiotics Promote Resistance

The selection process is fairly straightforward. When an antibiotic attacks a group of bacteria, cells that are highly susceptible to the medicine will die. But cells that have some resistance from the start, or that acquire it later (through mutation or gene exchange), may survive, especially if too little drug is given to overwhelm the cells that are present. Those cells, facing reduced competition from susceptible bacteria, will then go on to proliferate. When confronted with an antibiotic, the most resistant cells in a group will inevitably outcompete all others.

Promoting resistance in known pathogens is not the only self-defeating activity of antibiotics. When the medicines attack disease-causing bacteria, they also affect benign bacteria—innocent bystanders—in their path. They eliminate drug-susceptible bystanders that could otherwise limit the expansion of pathogens, and they simultaneously encourage the growth of resistant bystanders. Propagation of these resistant, nonpathogenic bacteria increases the reservoir of resistance traits in the bacterial population as a whole and raises the odds that such traits will spread to pathogens. In addition, sometimes the growing populations of bystanders themselves become agents of disease.

Widespread use of cephalosporin antibiotics, for example, has promoted the proliferation of the once benign intestinal bacterium *E. faecalis,* which is naturally resistant to those drugs. In most people, the immune system is able to check the growth of even multidrug-resistant *E. faecalis,* so that it does not produce illness. But in hospitalized patients with compromised immunity, the enterococcus can spread to the heart valves and other organs and establish deadly systemic disease.

Moreover, administration of vancomycin over the years has turned *E. faecalis* into a dangerous reservoir of vancomycin-resistance traits. Recall that some strains of the pathogen *S. aureus*



SPREAD OF RESISTANT BACTERIA, which occurs readily, can extend quite far. In one example, investigators traced a strain of multidrug-resistant *Streptococcus pneumoniae* from Spain to Portugal, France, Poland, the U.K., South Africa, the U.S. and Mexico.

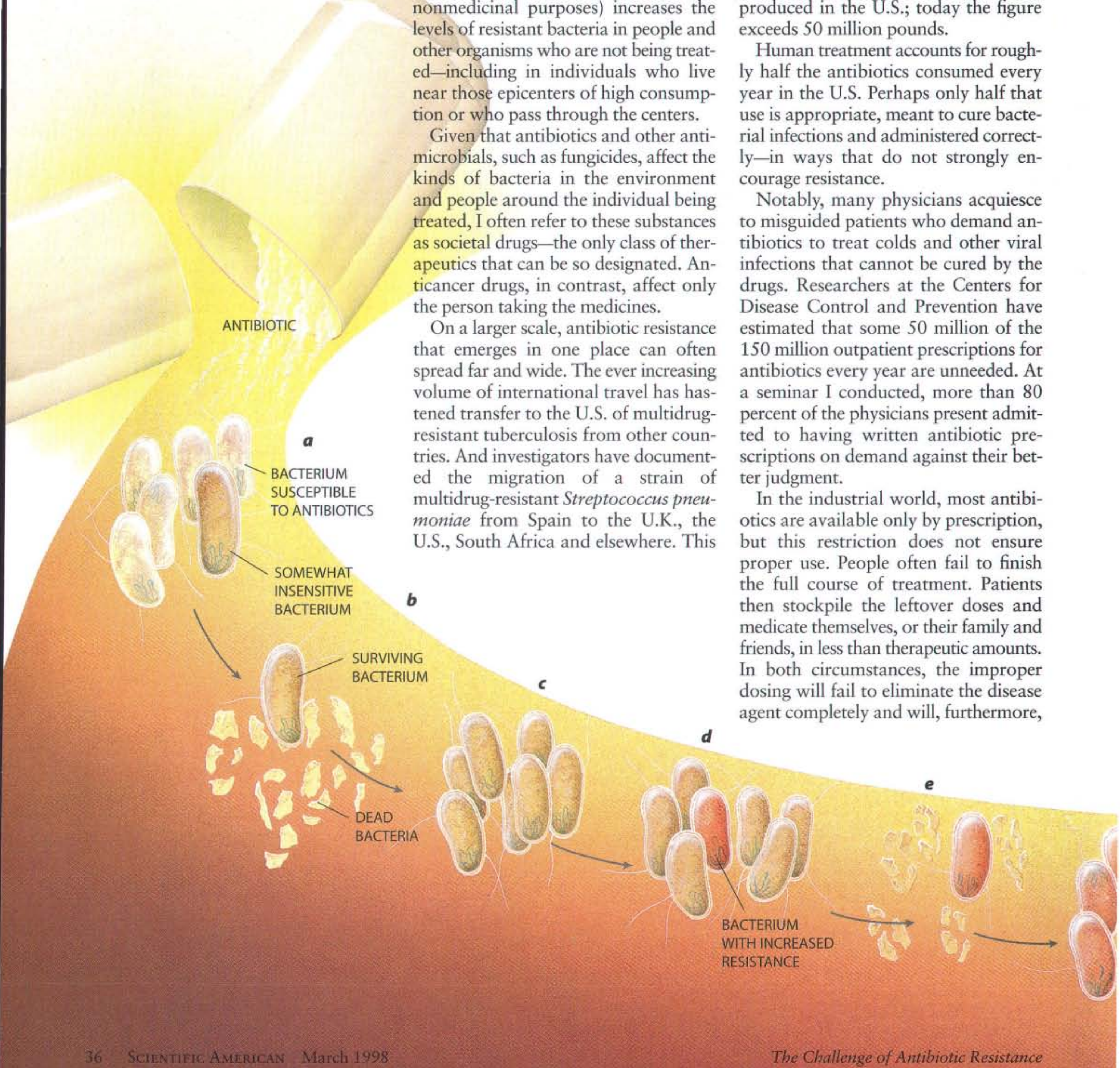
are multidrug-resistant and are responsive only to vancomycin. Because vancomycin-resistant *E. faecalis* has become quite common, public health experts fear that it will soon deliver strong vancomycin resistance to those *S. aureus* strains, making them incurable.

The bystander effect has also enabled multidrug-resistant strains of *Acinetobacter* and *Xanthomonas* to emerge and become agents of potentially fatal bloodborne infections in hospitalized patients. These formerly innocuous microbes were virtually unheard of just five years ago.

As I noted earlier, antibiotics affect the mix of resistant and nonresistant bacteria both in the individual being treated and in the environment. When resistant bacteria arise in treated individuals, these microbes, like other bacteria, spread readily to the surrounds and to new hosts. Investigators have shown that when one member of a household chronically takes an antibiotic to treat acne, the concentration of antibiotic-resistant bacteria on the skin of family members rises. Similarly, heavy use of antibiotics in such settings as hospitals, day care centers and farms (where the drugs are often given to livestock for nonmedicinal purposes) increases the levels of resistant bacteria in people and other organisms who are not being treated—including in individuals who live near those epicenters of high consumption or who pass through the centers.

Given that antibiotics and other antimicrobials, such as fungicides, affect the kinds of bacteria in the environment and people around the individual being treated, I often refer to these substances as societal drugs—the only class of therapeutics that can be so designated. Anticancer drugs, in contrast, affect only the person taking the medicines.

On a larger scale, antibiotic resistance that emerges in one place can often spread far and wide. The ever increasing volume of international travel has hastened transfer to the U.S. of multidrug-resistant tuberculosis from other countries. And investigators have documented the migration of a strain of multidrug-resistant *Streptococcus pneumoniae* from Spain to the U.K., the U.S., South Africa and elsewhere. This bacterium, also known as the pneumococcus, is a cause of pneumonia and meningitis, among other diseases.



Antibiotic Use Is Out of Control

For those who understand that antibiotic delivery selects for resistance, it is not surprising that the international community currently faces a major public health crisis. Antibiotic use (and misuse) has soared since the first commercial versions were introduced and now includes many nonmedicinal applications. In 1954 two million pounds were produced in the U.S.; today the figure exceeds 50 million pounds.

Human treatment accounts for roughly half the antibiotics consumed every year in the U.S. Perhaps only half that use is appropriate, meant to cure bacterial infections and administered correctly—in ways that do not strongly encourage resistance.

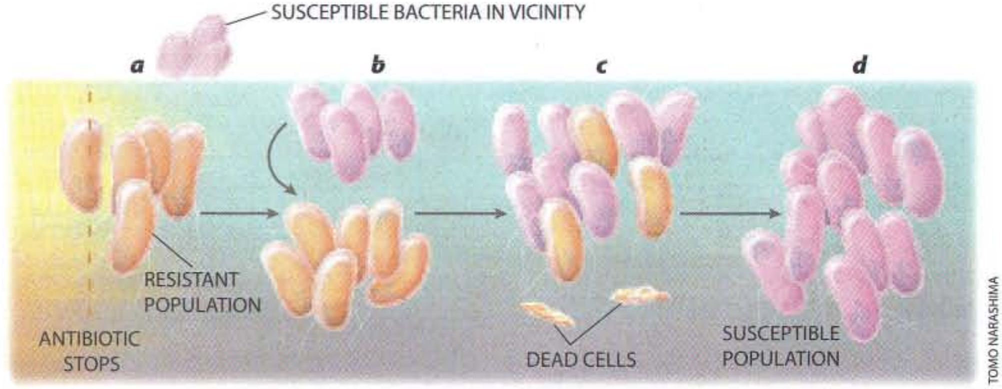
Notably, many physicians acquiesce to misguided patients who demand antibiotics to treat colds and other viral infections that cannot be cured by the drugs. Researchers at the Centers for Disease Control and Prevention have estimated that some 50 million of the 150 million outpatient prescriptions for antibiotics every year are unneeded. At a seminar I conducted, more than 80 percent of the physicians present admitted to having written antibiotic prescriptions on demand against their better judgment.

In the industrial world, most antibiotics are available only by prescription, but this restriction does not ensure proper use. People often fail to finish the full course of treatment. Patients then stockpile the leftover doses and medicate themselves, or their family and friends, in less than therapeutic amounts. In both circumstances, the improper dosing will fail to eliminate the disease agent completely and will, furthermore,

encourage growth of the most resistant strains, which may later produce hard-to-treat disorders.

In the developing world, antibiotic use is even less controlled. Many of the same drugs marketed in the industrial nations are available over the counter. Unfortunately, when resistance becomes a clinical problem, those countries, which often do not have access to expensive drugs, may have no substitutes available.

The same drugs prescribed for human therapy are widely exploited in animal husbandry and agriculture. More than 40 percent of the antibiotics manufactured in the U.S. are given to animals. Some of that amount goes to treating or preventing infection, but the lion's share is mixed into feed to promote growth. In this last application, amounts too small to combat infection are delivered for weeks or months at a time. No one is entirely sure how the drugs support growth. Clearly, though, this long-term exposure to low doses is the perfect formula for selecting increasing numbers of resistant bacteria in the treated animals—which may then pass the microbes to caretakers and, more broadly, to people who prepare and consume undercooked meat.



RESISTANT POPULATION of bacteria will disappear naturally only if susceptible bacteria live in the vicinity. After antibiotic therapy stops (*a*), resistant bacteria can persist for a while. If susceptible bacteria are nearby, however, they may recolonize the individual (*b*). In the absence of the drug, the susceptible bugs will have a slight survival advantage because they do not have to expend energy maintaining resistance genes. After a time, then, they may outcompete the resistant microbes (*c* and *d*). For this reason, protecting susceptible bacteria needs to be a public health priority.

In agriculture, antibiotics are applied as aerosols to acres of fruit trees, for controlling or preventing bacterial infections. High concentrations may kill all the bacteria on the trees at the time of spraying, but lingering antibiotic residues can encourage the growth of resistant bacteria that later colonize the fruit during processing and shipping. The aerosols also hit more than the targeted trees. They can be carried considerable distances to other trees and food plants, where they are too dilute to eliminate full-blown infections but are still capable of killing off sensitive bacteria and thus giving the edge to resistant versions. Here, again, resistant bacteria can make their way into people through the food chain, finding a home in the intestinal tract after the produce is eaten.

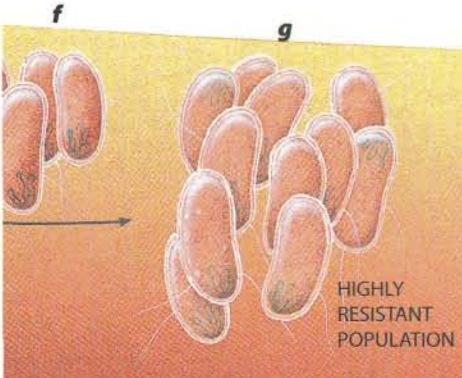
The amount of resistant bacteria people acquire from food apparently is not trivial. Denis E. Corpet of the National Institute for Agricultural Research in Toulouse, France, showed that when human volunteers went on a diet consisting only of bacteria-free foods, the number of resistant bacteria in their feces decreased 1,000-fold. This finding suggests that we deliver a supply of resistant strains to our intestinal tract whenever we eat raw or undercooked items. These bacteria usually are not harmful, but they could be if by chance a disease-causing type contaminated the food.

The extensive worldwide exploitation of antibiotics in medicine, animal care and agriculture constantly selects for strains of bacteria that are resistant to the drugs. Must all antibiotic use be halted to stem the rise of intractable bacteria? Certainly not. But if the drugs are to retain their power over pathogens, they have to be used more responsibly. Society can accept some increase in the fraction of resistant bacteria when a disease needs to be treated; the rise is unacceptable when antibiotic use is not essential.

Reversing Resistance

A number of corrective measures can be taken right now. As a start, farmers should be helped to find inexpensive alternatives for encouraging animal growth and protecting fruit trees. Improved hygiene, for instance, could go a long way to enhancing livestock development.

The public can wash raw fruit and vegetables thoroughly to clear off both resistant bacteria and possible antibiotic residues. When they receive prescriptions for antibiotics, they should complete the full course of therapy (to ensure that all the pathogenic bacteria die) and should not "save" any pills for later use. Consumers also should refrain from demanding antibiotics for colds and other viral infections and might consider seeking nonantibiotic therapies for minor conditions, such as certain cases of acne. They can continue to put antibiotic ointments on small cuts, but they should think twice about routinely us-



ANTIBIOTIC USE SELECTS—promotes the evolution and growth of—bacteria that are insensitive to that drug. When bacteria are exposed to an antibiotic (*a*), bacterial cells that are susceptible to the drug will die (*b*), but those with some insensitivity may survive and grow (*c*) if the amount of drug delivered is too low to eliminate every last cell. As treatment continues, some of the survivors are likely to acquire even stronger resistance (*d*)—either through a genetic mutation that generates a new resistance trait or through gene exchange with newly arriving bacteria. These resistant cells will then evade the drug most successfully (*e*) and will come to predominate (*f* and *g*).

ing hand lotions and a proliferation of other products now imbued with antibacterial agents. New laboratory findings indicate that certain of the bacteria-fighting chemicals being incorporated into consumer products can select for bacteria resistant both to the antibacterial preparations and to antibiotic drugs [*see box on page 34*].

Physicians, for their part, can take some immediate steps to minimize any resistance ensuing from required uses of antibiotics. When possible, they should try to identify the causative pathogen before beginning therapy, so they can prescribe an antibiotic targeted specifically to that microbe instead of having to choose a broad-spectrum product. Washing hands after seeing each patient is a major and obvious, but too often overlooked, precaution.

To avoid spreading multidrug-resistant infections between hospitalized patients, hospitals place the affected patients in separate rooms, where they are seen by gloved and gowned health workers and visitors. This practice should continue.

Having new antibiotics could provide more options for treatment. In the 1980s pharmaceutical manufacturers, thinking infectious diseases were essentially conquered, cut back severely on searching for additional antibiotics. At the time, if one drug failed, another in the arsenal would usually work (at least in the industrial nations, where supplies are plentiful). Now that this happy state of affairs is coming to an end, researchers are searching for novel antibiotics again. Regrettably, though, few drugs are likely to pass soon all technical and regulatory hurdles needed to reach the market. Furthermore, those that are close to being ready are structurally similar to existing antibiotics; they could easily encounter bacteria that already have defenses against them.

With such concerns in mind, scientists are also working on strategies that will give new life to existing antibiotics. Many bacteria evade penicillin and its relatives by switching on an enzyme, penicillinase, that degrades those compounds. An antidote already on pharmacy shelves contains an inhibitor of penicillinase; it prevents the breakdown of penicillin and so frees the antibiotic to work normally. In one of the strategies under study, my laboratory at Tufts University is developing a compound to jam a microbial pump that ejects tetracycline from bacteria; with the pump inactivated, tetracycline can penetrate bacterial cells effectively.

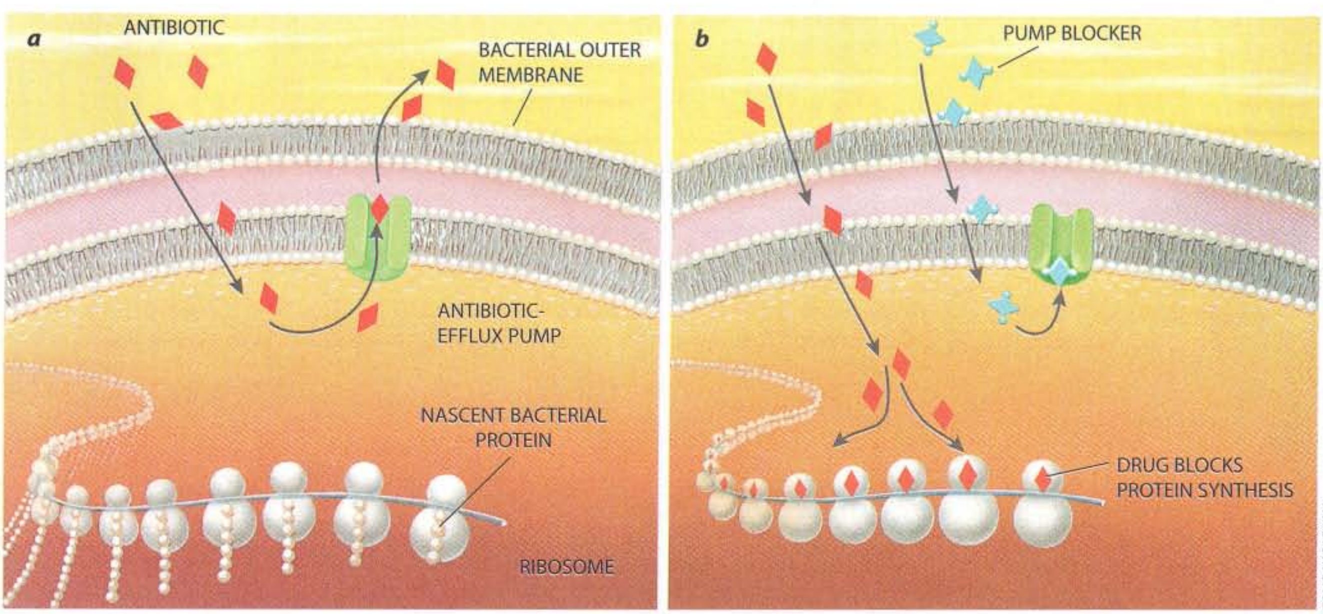
Considering the Environmental Impact

As exciting as the pharmaceutical research is, overall reversal of the bacterial resistance problem will require public health officials, physicians, farmers and others to think about the effects of antibiotics in new ways. Each time an antibiotic is delivered, the fraction of resistant bacteria in the treated individual and, potentially, in others, increases. These resistant strains endure for some time—often for weeks—after the drug is removed.

The main way resistant strains disappear is by squaring off with susceptible versions that persist in—or enter—a treated person after antibiotic use has stopped. In the absence of antibiotics, susceptible strains have a slight survival advantage, because the resistant bacteria have to divert some of their valuable energy from reproduction to maintaining antibiotic-fighting traits. Ultimately, the susceptible microbes will win out, if they are available in the first place and are not hit by more of the drug before they can prevail.

Correcting a resistance problem, then, requires both improved management of antibiotic use and restoration of the environmental bacteria susceptible to these drugs. If all reservoirs of susceptible bacteria were eliminated, resistant forms would face no competition for survival and would persist indefinitely.

In the ideal world, public health officials would know the extent of antibiotic resistance in both the infectious and benign bacteria in a community. To treat a specific pathogen, physicians would favor an antibiotic most likely to encounter little resistance from any bacteria in the community. And they would deliver enough antibiotic to clear the in-



ONE PHARMACEUTICAL STRATEGY for overcoming resistance capitalizes on the discovery that some bacteria defeat certain antibiotics, such as tetracycline, by pumping out the drugs (*a*). To combat that ploy, investigators are devising compounds that would jam the pumps (*b*), thereby freeing the antibiotics to function effectively. In the case of tetracycline, the antibiotic works by interfering with the ribosomes that manufacture bacterial proteins.

Some Actions Physicians and Consumers Can Take to Limit Resistance

The easy accessibility to antibiotics parodied in the cartoon is a big contributor to antibiotic resistance. This list suggests some immediate steps that can help control the problem. —*S.B.L.*

Physicians

- Wash hands thoroughly between patient visits.
- Do not accede to patients' demands for unneeded antibiotics.
- When possible, prescribe antibiotics that target only a narrow range of bacteria.
- Isolate hospital patients with multidrug-resistant infections.
- Familiarize yourself with local data on antibiotic resistance.

Consumers

- Do not demand antibiotics.
- When given antibiotics, take them exactly as prescribed and complete the full course of treatment; do not hoard pills for later use.
- Wash fruits and vegetables thoroughly; avoid raw eggs and undercooked meat, especially in ground form.
- Use soaps and other products with antibacterial chemicals only when protecting a sick person whose defenses are weakened.

"Don't forget to take a handful of our complimentary antibiotics on your way out."

fection completely but would not prolong therapy so much as to destroy all susceptible bystanders in the body.

Prescribers would also take into account the number of other individuals in the setting who are being treated with the same antibiotic. If many patients in a hospital ward were being given a particular antibiotic, this high density of use would strongly select for bacterial strains unsubmissive to that drug and would eliminate susceptible strains. The ecological effect on the ward would be broader than if the total amount of the antibiotic were divided among just a few people. If physicians considered the effects beyond their individual patients, they might decide to prescribe different antibiotics for different patients, or in different wards, thereby minimizing the selective force for resistance to a single medication.

Put another way, prescribers and public health officials might envision an "antibiotic threshold": a level of antibiotic usage able to correct the infections within a hospital or community but still falling below a threshold level that would strongly encourage propagation of resistant strains or would eliminate large numbers of competing, susceptible microbes. Keeping treatment levels below the threshold would ensure that the original microbial flora in a person or a community could be restored rapidly by susceptible bacteria in the vicinity after treatment ceased.

The problem, of course, is that no one yet knows how to determine where that threshold lies, and most hospitals and communities lack detailed data on the nature of their microbial populations. Yet with some dedicated work, researchers should be able to obtain both kinds of information.

Control of antibiotic resistance on a wider, international scale will require cooperation among countries around the globe and concerted efforts to educate the world's populations about drug resistance and the impact of improper antibiotic use. As a step in this direction, various groups are now attempting to track the emergence of resistant bacterial strains. For example, an international organization, the Alliance for the Prudent Use of Antibiotics (P.O. Box 1372, Boston, MA 02117), has been monitoring the worldwide emergence of such strains since 1981. The group shares information with members in more than 90 countries. It also produces educational brochures for the public and for health professionals.

The time has come for global society to accept bacteria as normal, generally beneficial components of the world and not try to eliminate them—except when they give rise to disease. Reversal of resistance requires a new awareness of the broad consequences of antibiotic use—a perspective that concerns itself not only with curing bacterial disease at the moment but also with preserving microbial communities in the long run, so that bacteria susceptible to antibiotics will always be there to outcompete resistant strains. Similar enlightenment should influence the use of drugs to combat parasites, fungi and viruses. Now that consumption of those medicines has begun to rise dramatically, troubling resistance to these other microorganisms has begun to climb as well. SA

The Author

STUART B. LEVY is professor of molecular biology and microbiology, professor of medicine and director of the Center for Adaptation Genetics and Drug Resistance at the Tufts University School of Medicine. He is also president of the Alliance for the Prudent Use of Antibiotics and president-elect of the American Society for Microbiology.

Further Reading

THE ANTIBIOTIC PARADOX: HOW MIRACLE DRUGS ARE DESTROYING THE MIRACLE. S. B. Levy. Plenum Publishers, 1992.

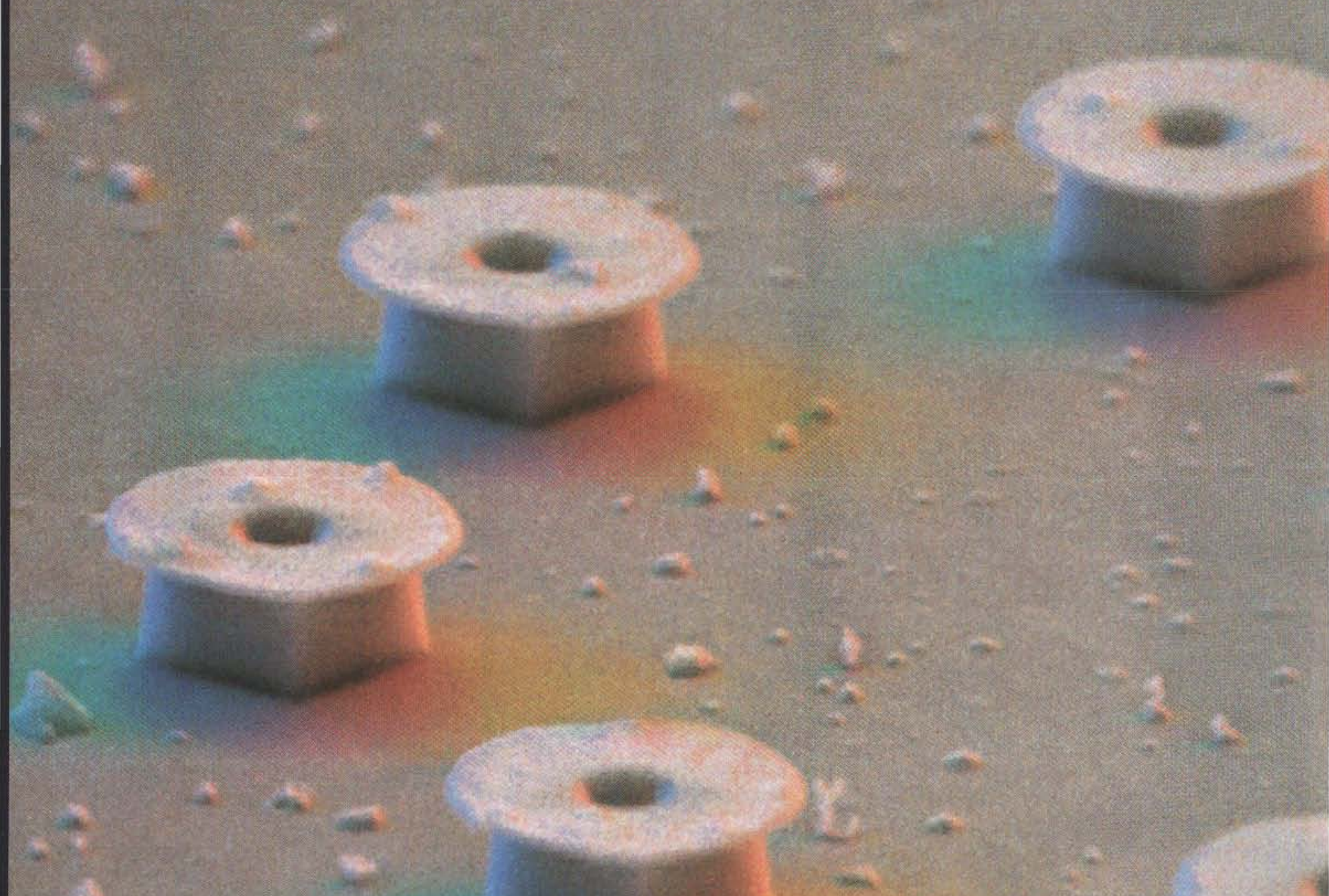
DRUG RESISTANCE: THE NEW APOCALYPSE. Special issue of *Trends in Microbiology,* Vol. 2, No. 10, pages 341–425; October 1, 1994.

ANTIBIOTIC RESISTANCE: ORIGINS, EVOLUTION, SELECTION AND SPREAD. Edited by D. J. Chadwick and J. Goode. John Wiley & Sons, 1997.

Nanolasers

Semiconductor lasers have shrunk to dimensions even smaller than the wavelength of the light they emit. In that realm, quantum behavior takes over, enabling more efficient and faster devices

by Paul L. Gourley



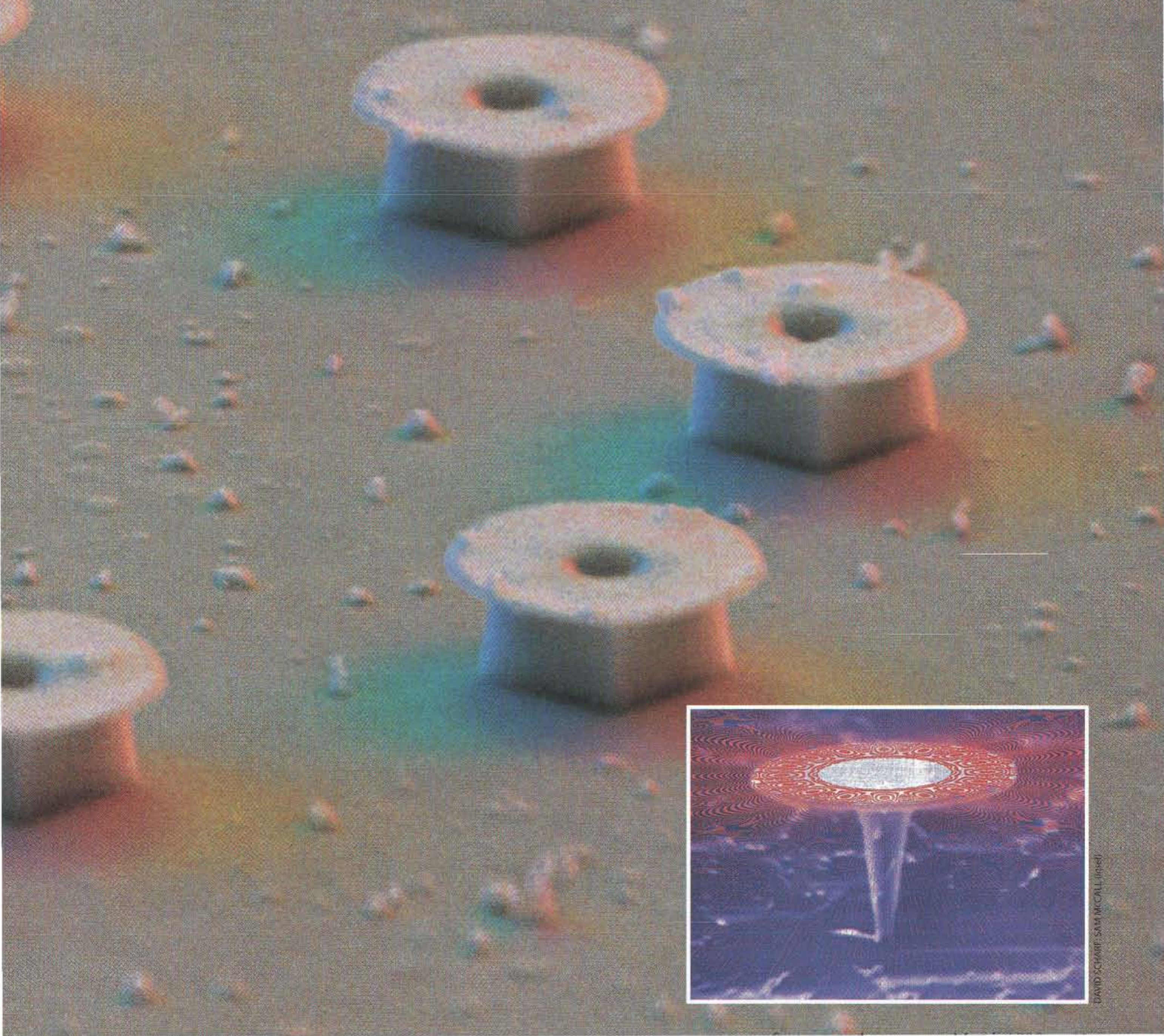
MICRODISK LASERS are each only a couple of microns, or millionths of a meter, in diameter and just a fraction of a micron thick. The disks are made of semiconductor material and are supported by pedestals. Light is generated within the disk and skims along its circumference before escaping radially, as shown by the red wave pattern in the computer simulation (*inset at right*). The depressions in the center of the disks and the tiny, random particles are artifacts of the chemical etching process used to fabricate the structures.

For decades, silicon transistors have become smaller and smaller, allowing the fabrication of tiny but powerful chips. Less well known is the parallel revolution of semiconductor lasers. Recently researchers have shrunk some of the dimensions of such devices to an astonishing scale of nanometers (billionths of meters), even smaller than the wavelength of the light they produce. At such sizes—less than one hundredth the thickness of a human hair—curious aspects of quantum physics begin to take over. By exploiting this quantum behavior, researchers can tailor the basic characteristics of the devices to achieve even greater efficiencies and faster speeds.

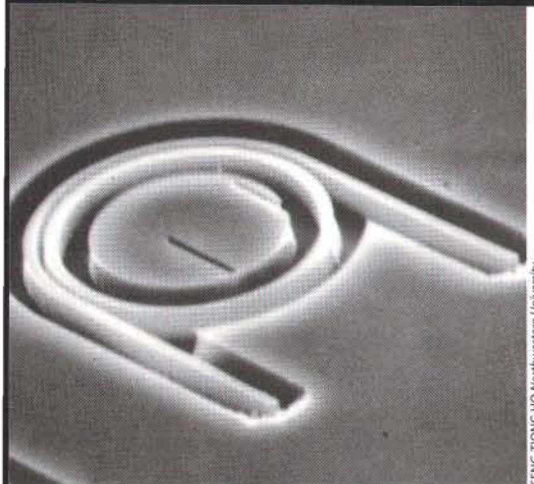
Nanolasers could have myriad applications, for instance, in optical computers, where light would replace electricity for transporting, processing and storing information. Even though light-based computing may not occur anytime soon, other uses, such as in fiber-optic communications, have now become increasingly practical. With other researchers, I am also investigating the new lasers for novel purposes, such as the early detection of disease.

Jumping Electrons

Although nanolasers push the boundaries of modern physics, the devices work much like their earliest ancestor, a contraption fashioned from a rod of dark ruby more than 35 years ago. Essentially, a lasing material—for example, a gas such as helium or neon, or a crystalline semiconductor—is sandwiched between two mirrors. The substance is "pumped" with light or electricity. The process excites the electrons in the material to hop from lower to higher energy levels. When the electrons return to the lower stations, they produce light,



DAVID SCHARF; SAM McCALL (inset)



SENG-TIONG HO *Northwestern University*

MICRORING LASER is surrounded by a U-shaped glass structure that guides the light out of the device in two parallel beams along the legs of the U. The laser is essentially an extremely thin semiconductor wire—with a rectangular cross section of 400 by 200 nanometers, or billionths of a meter—curled into the shape of an ultraskinny doughnut.

which is reflected between the mirrors.

The bouncing photons trigger other "excited" electrons—those in higher energy states—to emit identical photons, much like firecrackers that pop and set off other firecrackers. This chain reaction is called stimulated emission. (Hence the name "laser," which is an acronym for "*l*ight *a*mplification by *s*timulated *e*mission of *r*adiation.") As the number of photons grows, they become part of a communal wave that intensifies, finally bursting through one of the mirrors in a concentrated, focused beam.

But not all the photons take part in this wave. In fact, many are emitted spontaneously, apart from the chain reaction. In a large space—to a subatomic particle, the size of a typical laser cavity is immense—photons are relatively free to do what they want. Thus, many of the free-spirited photons are literally on a different wavelength, and they can scatter in all directions, often hitting the sides of the laser and generating unwanted heat instead of bouncing between the mirrors. For some types of lasers, only one photon in 10,000 is useful.

Because of this enormous waste, a certain threshold of energy is necessary to ensure that the number of excited electrons is large enough to induce and maintain stimulated emission. The requirement is analogous to the minimum amount of heat needed to bring a pot of water to boil. If the hurdle is not cleared, the laser will fail to attain the self-sustaining chain reaction crucial to its operation. This obstacle is why semiconductor lasers have required relatively high currents to work, in contrast to silicon transistors, which are much more frugal. But if semiconductor lasers could stop squandering energy, they could become competitive with their electronic counterparts for a host of applications, including their use in computers.

Recently the concept of "thresholdless" operation has become increasingly favored by many physicists. Proposed by Yoshihisa Yamamoto of NTT Basic Research Laboratories and Stanford University and Takeshi Kobayashi of Osaka University in Japan, thresholdless operation calls for all photons, even those spontaneously born, to be drafted into lasing duty. In theory, the device would require only the tiniest amount of energy, almost like a special kettle that could boil water with the heat of just a single match. Researchers disagree about the best design of such a laser. The consensus, though, is that the dimensions must be extraordinarily small—on the order of the wavelength of light emitted—so that the devices could take advantage of quantum behavior.

A New Generation

The groundwork for thresholdless operation was set in the late 1970s, when Kenichi Iga and other researchers at the Tokyo Institute of Technology demonstrated a radically different type of semiconductor laser [see "Microlasers," by J. L. Jewell, J. P. Harbison and A. Scherer; SCIENTIFIC AMERICAN, November 1991]. Popularly referred to as microlasers because of their micron-size dimensions, these devices are cousins to the semiconductor diode lasers widely found in compact-disc players. ("Diode" refers to a one-way flow of electricity during operation.)

Microlasers, however, differ from their common diode relatives in several fundamental ways. The latter are shaped like rectangular boxes that must be cleaved, or diced, from a large wafer, and they issue light longitudinally from the cut edges. Microlasers are smaller, cylindrical shapes formed by etching, and they emit light from the top—perpendicular to the round layers of semiconductor material that make up the device. Therefore, microlasers produce more perfectly circular beams. In addition, they can be built and tested many at a time in arrays on a wafer, similar to the way in which computer chips are fabricated. In contrast, diode lasers must generally be tested individually after having been diced into separate units.

Perhaps more important, microlasers exploit the quantum behavior of both electrons and photons. The devices are built with a "well"—an extremely thin layer of semiconductor only several atoms thick. In such a minute space, electrons can exist only at certain discrete, or quantized, energy levels separated by forbidden territory, called the band gap of the semiconductor. By sandwiching the quantum well with other material, researchers can trap electrons and force them to jump across the band gap to emit just the right kind of light.

Microlasers must also imprison photons to function. To accomplish this feat, engineers take advantage of the same effect that causes a transparent window to display a faint reflection. This phenomenon results from glass having a higher refractive index than air—that is, photons move more slowly through glass. When light passes between materials with different refractive indices, some of the photons are reflected at the border. The mirrors of microlasers consist of alternating layers of semiconductors with different refractive indices (such as gallium arsenide and aluminum arsenide). If the layers are just one quarter of a wavelength thick, the geometry of the structure will allow the weak reflections to reinforce one another. For the coupling of gallium arsenide and aluminum arsenide, a dozen pairs of layers will bounce back 99 percent of the light—a performance superior to that of polished metal mirrors commonly found in bathrooms.

Already the first crop of microlasers has found commercial applications in fiber-optic communications. Other uses are currently under investigation [*see box on page 44*]. Meanwhile ongoing work continues to refine the structures. In one recent device, certain layers are selectively oxidized, which helps to raise the population of excited electrons and bouncing photons in the well area, resulting in an operating efficiency greater than 50 percent. In other words, the laser is able to convert more than half the input energy into output laser light. This performance far exceeds that of semiconductor diode lasers, which are typically not even 30 percent efficient.

Microlasers have led to a new generation of devices that exploits electronic quantum behavior further. Scientists have now built structures such as quantum wires and dots that confine electrons to one and zero dimensions, re-

spectively. (Wells restrict them to two.) Additionally, in a fundamentally new device called the quantum-cascade laser, researchers at Bell Laboratories have strung together many quantum wells, like a series of small waterfalls. In such a laser, an electron returning to a lower energy state will not take one big bandgap jump but multiple smaller ones, emitting a photon at each successive hop—thereby increasing the lasing chain reaction. An exciting feature of this innovative laser is that it allows engineers to tailor the type of light produced by adjusting the width of the wells; therefore, the electronic band gap of the material—a property ordained by nature—no longer dictates the kind of photons produced.

In a separate but related track of research, scientists have been exploring quantum-optical behavior. To do so, investigators have had to shrink some of the dimensions of the devices to smaller than even the wavelength of the light emitted. In that microscopic world, photons are restricted to certain discrete states, similar to the restraints placed on electrons trapped in quantum wells.

A Short Guitar String

Large lasers emit various types of photons, just as a long guitar string, when strummed, produces a sound consisting of a fundamental frequency (corresponding to the pitch) and many overtones. But as the guitar string is made shorter, the pitch becomes higher and the number of overtones decreases until the process reaches a limit decreed by the thickness and type of material of the string.

Similarly, physicists have been shrinking lasers to restrict the number of states, or modes, that the photons can inhabit. A limit to this miniaturization is one half the wavelength of the light emitted, because this dimension is the smallest for which the light is able to bounce between the mirrors. At this minimum boundary, photons would have just one possible state, corresponding to the fundamental optical mode of the device. Because of this Hobson's choice, every photon would be forced to contribute to the communal wave (the fundamental mode) that intensifies into the beam of light that finally bursts through one of the mirrors. In other words, no photons would go to waste: the laser would be thresholdless.

With colleagues at Sandia National Laboratories, I observed such quantized photon states in experiments more than a decade ago. By bringing the end mirrors of a microlaser closer, we were able to squeeze the broad spectrum of photons emitted into just a few optical modes. We showed that these modes occurred at wavelengths whose integral multiples were equal to the round-trip distance between the mirrors, in the same way that a guitar string can vibrate with four or five wavelengths between its fixed ends but not with four and one-sixth wavelengths. Furthermore, we verified that we could enhance these effects by moving the mirrors closer, approaching the limit of one half wavelength (hundreds of nanometers). But these devices were not yet thresholdless. Even the most advanced microlasers, which might now be legitimately called nanolasers, allow about 100 photonic states—much improved from the tens of thousands of options available to photons in conventional diode lasers but still not acceptable for entrance into thresholdless nirvana.

To achieve that ideal, researchers have recently begun to investigate other nanometer-scale geometries. One such design is the microdisk laser, developed by Richart E. Slusher and his colleagues at Bell Labs. With advanced etching processes similar to those used to fabricate computer chips, the Bell Labs researchers have been able to carve an ultrathin disk a couple of microns in diameter and just 100 nanometers thick. The semiconductor disk is surrounded by air and supported by a tiny pedestal, making the overall structure look like a microscopic table.

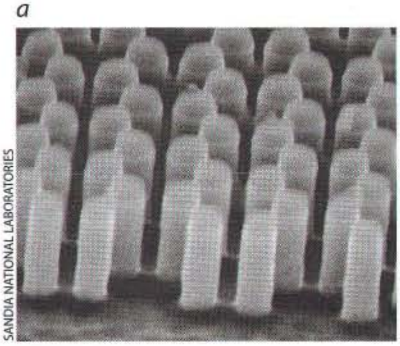
Because the semiconductor and air have very different indices of refraction, light generated inside the disk reflects

Photonic Lattices Block Light

Tiny gallium arsenide posts will block infrared light when the individual columns are arranged in a hexagonal lattice of just the right spacing (*a*). The periodicity of the structure, combined with the difference in the speed of light through the semiconductor posts and the surrounding air, results in multiple refractions and reflections that effectively block light over a range of wavelengths, as shown in a light-scattering micrograph (*b*) of a similar lattice (*inset in b*).

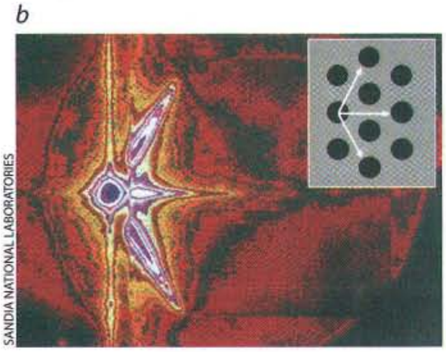
The concept also works in one dimension, as demonstrated by a semiconductor bridge punched lengthwise with holes (*c*). Light traveling across the bridge is blocked by the one-dimensional "array" of holes, which are analogous to the posts in the hexagonal lattice. By purposely introducing a "defect"—the slightly larger spacing between the two holes in the center of the bridge—researchers can change the reflection and refraction pattern within the structure. The irregular spacing circumscribes a minuscule "box," with a volume of only a twentieth of a cubic micron, that could be developed into a laser. —*P.L.G.*

a



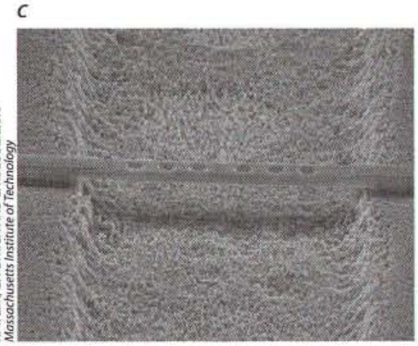
SANDIA NATIONAL LABORATORIES

b



SANDIA NATIONAL LABORATORIES

c



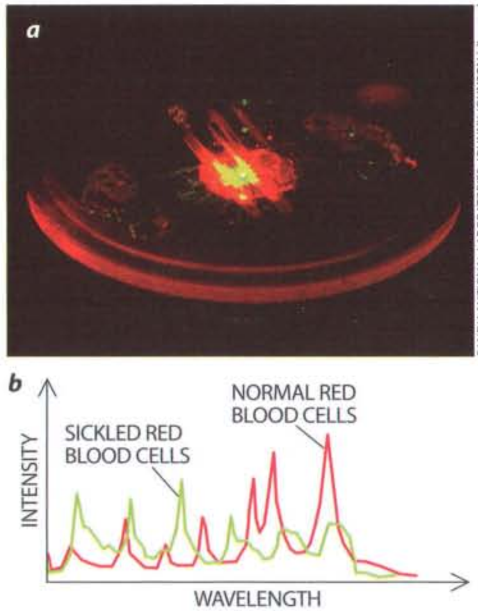
K.-Y. LIM, G. PETRICH AND L. KOLODZIEJSKI *Massachusetts Institute of Technology*

Biocavity Lasers for Disease Detection

As semiconductor lasers continue getting smaller, faster and more efficient, they will enable an increasing number of novel applications. One possibility is the detection of disease. At Sandia National Laboratories, my colleagues and I have developed a "biocavity laser" (*a*), which can, for example, be used to distinguish cancerous cells from normal ones.

The device is basically a microlaser—a tiny piece of gallium arsenide sandwiched between two mirrors. Infrared light that the semiconductor compound emits will repeatedly bounce between the mirrors, intensifying until it finally bursts out of the structure in a concentrated laser beam. To build a biocavity laser, we placed a thin layer of human tissue between the gallium arsenide and one of the mirrors. The organic material becomes part of the device itself, acting as an internal lens to focus the light. Thus, the size, shape and composition of the cells alter the laser beam by introducing overtones that result in a unique spectral signature. Doctors can use that information to distinguish between diseased and healthy tissue because the two types will result in different light spectra (*b*), just as a piccolo and flute playing the same note can be discerned by the distinct sound spectra of overtones produced by the two similar—yet unique—instruments.

Recently Anthony McDonald, Guild Copeland and I at Sandia worked with my brother Mark Gourley, an immunologist at the National Institutes of Health, to patent a portable, handheld version of the biocavity laser that doctors can use to analyze blood without having to send samples to a laboratory. In the device, blood flows through tiny grooves, each just a tenth the width of a human hair, that have been etched into one of the mirrors. By analyzing the resulting laser beam, the device can quickly detect the presence of crescent-shaped red blood cells—an indicator of sickle cell anemia. Doctors could also use the laser to study nanometer-scale changes in the cellular structure of blood that might be caused by the AIDS virus.



SANDIA NATIONAL LABORATORIES; JOHNNY JOHNSON (*diagram*)

In other experiments, biocavity lasers have also been able to discriminate between normal and cancerous cervical cells, as in Pap smears. Further advancements might even lead to a device for analyzing DNA.

The new technology boasts several advantages over conventional methods of tissue analysis, which require chemical staining to make cellular structures visible under microscopic examination in the lab. Such techniques rely heavily on qualitative human vision and are thus prone to error. In contrast, biocavity lasers produce simple, straightforward spectra that a handheld device can analyze almost instantly in clinics, offices and research laboratories, as well as in the field. *—P.L.G.*

within the structure, skimming along its circumference. The effect is similar to the "whispering gallery" sound waves first described by Lord Rayleigh more than a century ago. The physicist explained how conversations can be heard at opposite ends inside the great dome of St. Paul's Cathedral in London because the audible vibrations reflect off the walls and reinforce one another.

The tiny size of the microdisk restricts the photons to just a limited number of states, including the desired fundamental optical mode, while the whispering-gallery effect confines the photons until the light wave generated has built up enough energy to burst outside the structure. The result is extremely efficient operation with a low threshold. In fact, these microdisk lasers have worked with only about 100 microamps.

A variation of the microdisk is the microring laser, which is essentially a photonic wire curled into the shape of an ultraskinny doughnut. Seng-Tiong Ho and his colleagues at Northwestern University used microlithography to etch such a semiconductor structure with a diameter of 4.5 microns and a rectangular cross section measuring only 400 by 200 nanometers. To improve the quality of the light emitted, the Northwestern researchers surrounded the microring with a U-shaped glass structure that guides the photons out in two parallel beams along the legs of the U.

These novel devices have proved how the size and shape of a nanolaser can affect its operation by controlling the quantum behavior of the photons emitted. Investigators have recently pushed the technology even further, shrinking photonic wires to an amazing volume of just one fifth of a cubic micron. At that dimension, the structure has fewer than 10 photonic states—which approaches the conditions required for thresholdless operation.

Although these new nanolasers have reduced the *types* of photons to quantum-mechanical levels, they have not decreased the *number* of photons to such limits. With a small enough population, the behavior of light can be fundamentally altered for useful purposes. In recent landmark work, researchers at the Massachusetts Institute of Technology have shown that single excited atoms of barium can be fed one by one into a laser, with each atom emitting a useful photon. This incredibly efficient device is able to work with just 11 photons bouncing between the mirrors. Physicists are currently investigating such novel quantum optics for semiconductor nanolasers.

Stopping Light Periodically

A radically different approach to the design of nanolasers is to build a structure with materials that alternate at regular tiny intervals. If designed properly, the periodic modulation will imprison light by repeatedly reflecting it within the structure. This concept was first deployed by scientists who engineered the layered mirrors of micro-

lasers, which contain light in one dimension. Eli Yablonovitch, now at the University of California at Los Angeles, and researchers at the Department of Energy Ames Laboratory at Iowa State University extended the principle into two and three dimensions by proposing new structures called photonic lattices.

The overall concept is based on a phenomenon observed in the early 1900s by the father-and-son team of William Henry Bragg and William Lawrence Bragg. The English physicists, who shared a Nobel Prize in 1915, studied how x-rays striking a crystal will backscatter in a manner dependent on the periodic structure of the crystal lattice. In what is now known as Bragg's law, the two scientists stated that the intensity of reflected radiation depends on three factors: the wavelength of the x-rays, the spacing of the atoms in the crystal and the angle at which the x-rays strike the lattice.

Applying this knowledge to optical frequencies, investigators such as Thomas F. Krauss and Richard M. De La Rue of the University of Glasgow have shown that a lattice of two different alternating materials will backscatter light in a similar way. Furthermore, by using materials of very different indices of refraction and by selecting the right periodic spacing between those substances, researchers have shown that they can tailor and extend the range of wavelengths that the device reflects, in effect creating a "photonic band gap" similar to the forbidden territory of electrons in semiconductors.

At Sandia, Joel Wendt, Allen Vawter and I fabricated such a structure by building a hexagonal lattice of gallium arsenide posts—a design that was developed by John D. Joannopoulos and other researchers at M.I.T. By taking into account the different indices of refraction of gallium arsenide and the surrounding air, we determined the exact spacing of the posts necessary to trap infrared light.

Although we have demonstrated the feasibility of confining light in this two-dimensional array, we have not yet been able to turn the structure into a laser. One possible way to do so would be to pump one of the posts, making it emit light, which would then be repeatedly reflected (and effectively contained) by the other posts in the array. Basically, the lattice would act like the parallel mirrors in a traditional laser.

Using a reverse design in which the "posts" are made of air and the surrounding material is a semiconductor, the M.I.T. researchers have fabricated a tiny silicon bridge (470 nanometers wide and 200 nanometers thick) etched lengthwise with a single row of microscopic holes. Light is confined to traveling across the structure because of the difference in indices of refraction between the semiconductor and the surrounding air.

The M.I.T. scientists, including Joannopoulos, Pierre R. Villeneuve and Shanhui Fan, used computer simulations to determine the precise periodic spacing of the holes in order to define a one-dimensional array suitable for confining infrared light. Furthermore, the researchers introduced a "defect" into the lattice by making the distance between two adjacent holes near the center of the strip slightly larger. The irregularity creates a fundamental optical mode within the minuscule volume circumscribed by the nonuniform spacing. This "box" might one day be developed into a laser cavity, with the adjacent holes acting like mirrors. Amazingly, the box is a mere twentieth of a cubic micron. The M.I.T. group has since refined the structure, building it on a glass base, and has verified the computer simulations with experimental results.

Other work has investigated photonic lattices that vary periodically in three dimensions. But such structures have been difficult to build because microfabrication techniques such as electron-beam lithography are more suited for the two-dimensional patterning of chipmaking. Still, three-dimensional photonic lattices would theoretically confine light in all directions—an ideal feature for a thresholdless laser.

Whither Optical Computers?

In addition to higher efficiency, thresholdless operation could lead to ultrafast devices, which can be switched on and off instantaneously because they require such little energy for lasing to occur. In other words, waiting for a pot of water to boil can be brief if just one match will do the trick. Already some lasers can be switched on and off faster than 20 billion times a second.

Such blinding speeds are a natural for fiber-optic communications. Other applications will arise as these devices continue to become even faster, smaller and more energy-efficient. Thresholdless lasers, now a distinct possibility because of recent advances in the fabrication of structures of nanometer scale, hold great promise as components for transmitting, storing and manipulating information—that is, as the crucial building blocks for an optical computer. Ironically, advances in the shrinking of silicon transistors have enabled substantial improvements in semiconductor lasers, which could one day power computers and replace those tiny electronic circuits with optical ones. SA

The Author

PAUL L. GOURLEY is currently with Sandia National Laboratories, where he has been a principal investigator in the laboratory's fundamental work in semiconductor vertical-cavity, surface-emitting lasers. He earned an undergraduate degree in physics from the University of North Dakota and M.S. and Ph.D. degrees in physics from the University of Illinois. A fellow of the American Physical Society and the Optical Society of America, Gourley has received awards from *R&D* magazine for his invention of the biocavity laser and from the Department of Energy for his work in semiconductor photonics. His research interests include the study of semiconductor devices such as nanolasers and their various applications, particularly those in the biomedical field.

Further Reading

PHYSICS AND DEVICE APPLICATIONS OF OPTICAL MICROCAVITIES. H. Yokoyama in *Science,* Vol. 256, pages 66–70; April 3, 1992.

OPTICS OF NANOSTRUCTURES. Special issue of *Physics Today,* Vol. 46, No. 6; June 1993.

SURFACE-EMITTING LASERS. Paul L. Gourley, Kevin Lear and Richard Schneider, Jr., in *IEEE Spectrum,* Vol. 31, No. 8, pages 31–37; August 1994.

MICROSTRUCTURED SEMICONDUCTOR LASERS FOR HIGH-SPEED INFORMATION PROCESSING. Paul L. Gourley in *Nature,* Vol. 371, pages 571–577; October 13, 1994.

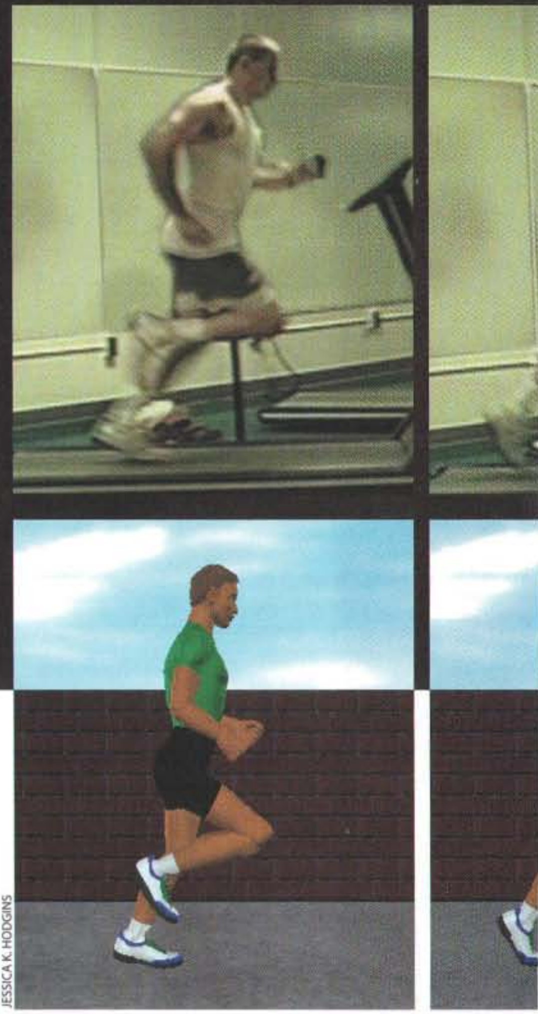
PHOTONIC CRYSTALS: MOLDING THE FLOW OF LIGHT. John D. Joannopoulos, Robert D. Meade and Joshua N. Winn. Princeton University Press, 1995.

MICROCAVITIES AND PHOTONIC BANDGAPS: PHYSICS AND APPLICATIONS. Edited by John Rarity and Claude Weisbuch. Kluwer Academic Publishers, 1996.

Animating Human Motion

Computer animation is becoming increasingly lifelike. Using simulation, a technique based on the laws of physics, researchers have created virtual humans who run, dive, bicycle and vault

by Jessica K. Hodgins



JESSICA K. HODGINS

People are skilled at perceiving the subtle details of human motion. A person can, for example, often recognize friends at a distance purely from their walk. Because of this ability, people have high standards for animations that feature humans. For computer-generated motion to be realistic and compelling, the virtual actors must move with a natural-looking style.

Synthetic human motion is needed for such applications as animation, virtual environments and video games. Animators would like to be able to create a *Toy Story* in which the children get as much screen time as their toys. Coaches could use virtual competitors to motivate and teach aspiring athletes. Video game designers could create products with highly interactive, engaging characters. The ability to simulate human motion also has significant scientific applications in ergonomics, gait analysis of athletes and physical rehabilitation.

Although the applications for synthetic human motion are many, specifying movement to a computer is surprisingly hard. Even a simple bouncing ball can be difficult to animate convincingly, in part because people quickly pick out action that is unnatural or implausible without necessarily knowing exactly what is wrong. Animation of a human is especially time-consuming because subtleties of the motion must be captured to convey personality and mood.

The techniques for computer animation fall into three basic categories: keyframing, motion capture and simulation. All three involve a trade-off between the level of control that the animator has over the fine details of the motion and the amount of work that the computer does on its own. Keyframing allows subtle control but requires that the animator ensure the naturalness of the result. Motion capture and simulation generate motion in a fairly automatic fashion but offer little opportunity for fine-tuning.

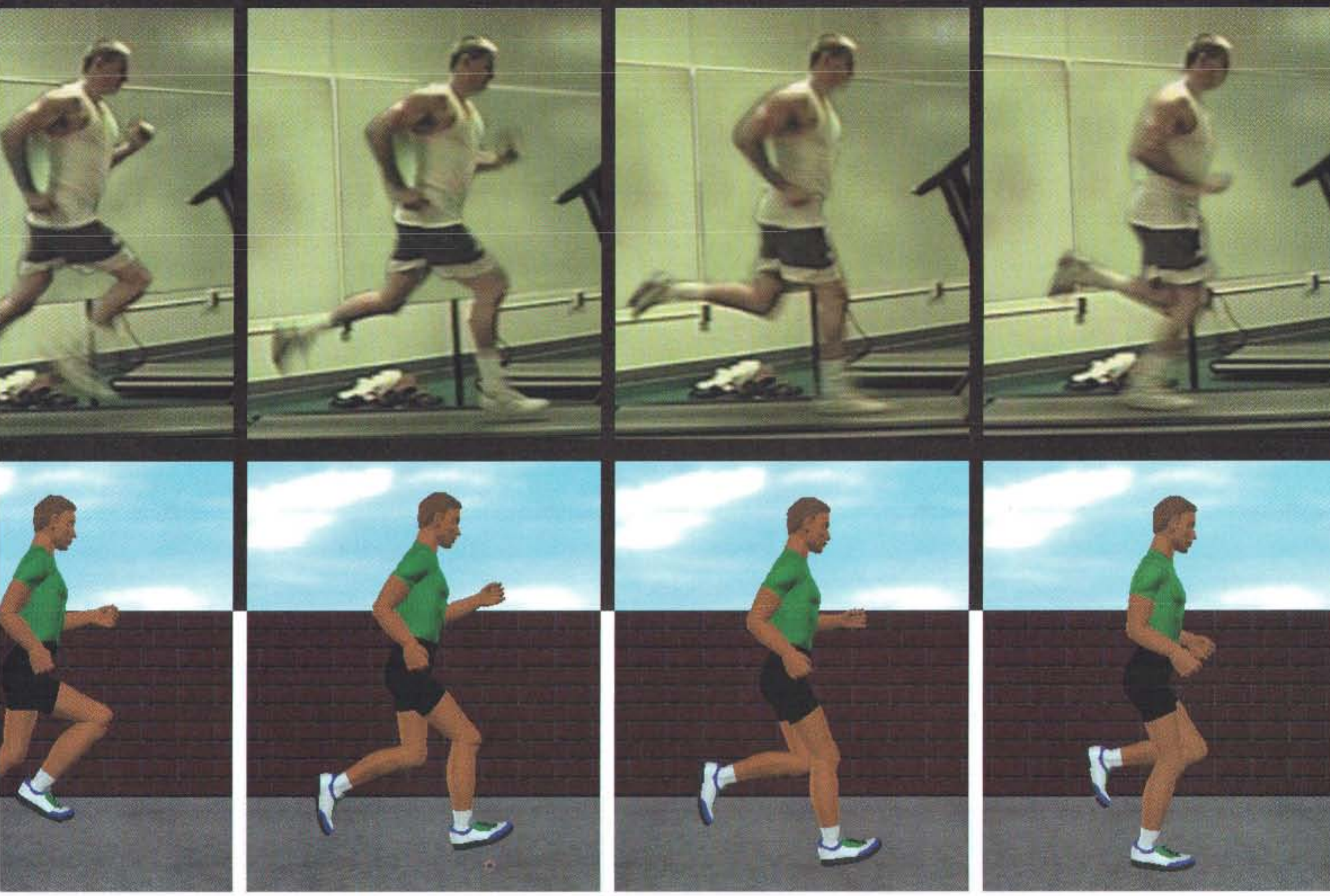
Keyframing and Motion Capture

Borrowing its name from a traditional hand animation technique, keyframing requires that the animator specify critical, or key, positions for the objects. The computer then fills in the missing frames by smoothly interpolating between those positions. The characters of the 1995 movie *Toy Story* were animated in this fashion, with more than 700 controls for each main character. For example, separate controls enabled movement of the different parts of a character's eyebrow.

The specification of keyframes can be partially automated with techniques that aid in the placement of some body joints. If the hand of a character must be in a particular location, for instance, the computer could calculate the appropriate elbow and shoulder angles. Although such techniques simplify the process, keyframing nonetheless requires that the animator possess a detailed understanding of how moving objects should behave over time as well as the talent to express that information through keyframed configurations. The continued popularity of keyframing comes from the degree of control that it allows over the subtle details of the motion.

Another technique is motion capture, in which magnetic or vision-based sensors record the actions of a human or animal subject in three dimensions. A computer then uses these data to animate a character. This technology has enabled a number of famous athletes to supply the actions for characters in sports video games.

Motion capture is growing in popularity because of the relative ease with which many commonplace human actions can be recorded; however, a number of problems prevent it from being an ideal solution for all applications. First, accurately measuring the motion of the human body is tricky because markers placed on skin and clothing shift as the performer moves, creating errors in the data. Second, discrepancies between the shapes or dimensions of the subject and the graphical character can lead to problems. If, for example, the subject was recorded touching a real table, the hands



of a shorter graphical character might appear sunk into the table.

Finally, the current technology makes it difficult to record certain movements. Magnetic systems often require the subject to be connected to a computer by cables, restricting the range of motion. These systems also produce noisy data when metal objects or equipment such as treadmills, which might be used to capture the motion of the subject running, are close by. Optical systems have problems with occlusion caused by one body part blocking another from view. In spite of these drawbacks, much of the motion in commercial animation is generated by modifying captured data to match the size and desired behavior of a virtual character.

Simulation

Unlike keyframing and motion capture, simulation uses the laws of physics to generate motion of figures and other objects. Virtual humans are usually represented as a collection of rigid body parts. The lower body, for instance, might consist of a torso, upper and lower legs, and feet connected by rotary joints for the hips, knees and ankles. My students and I have used biomechanical data from various studies, including measurements from cadavers, to build accurate models. The forearm of our virtual adult male, for example, has a mass of 1.1 kilograms (a weight of 2.4 pounds), a length of 0.24 meter (9.4 inches) and an average circumference of roughly 0.25 meter.

Although the models are physically plausible, they are nonetheless only an approximation of the human body. A collection of rigid body parts ignores the movement of muscle mass relative to bone, and although the shoulder is often modeled as a single joint with three degrees of freedom, the human clavicle and scapula allow more complex motions, such as shrugging. Recently researchers have begun to build more complex models, and the resulting simulations will become increasingly lifelike as researchers continue to add such detail.

When the models are of inanimate objects, such as clothing or water, the computer can determine their movements by making them obey equations of motion derived from physical laws. In the case of a ball rolling down a hill, the simulation could calculate motion by taking into account gravity and forces such as friction that result from the contact between the ball and the ground. But people have internal sources of energy and are not merely passive or inanimate objects. Virtual humans, therefore, require a source of muscle or motor commands—a "control system." This software computes and applies torques at each joint of the simulated body to enable the character to perform the desired action. A control system for jogging, for instance, must determine the torques necessary to swing the leg for-

RUNNING can be simulated by using the laws of physics to generate the motion. To create the sequence above (*second row*), a computer determined the torques necessary to swing the right leg forward before touchdown to prevent the runner from tripping. Using numerous such calculations, computers can synthesize the mechanics of how people run (*top row*). The images in both the real and simulated footage are spaced 0.066 second apart.

Three Animation Techniques

Keyframing

Toy Story relied heavily on keyframing, in which the animator specifies key positions for objects. A computer then fills in the missing frames by automatically interpolating between those positions.

EVERETT COLLECTION

ward before touchdown to prevent the runner from tripping.

My students and I have developed control systems for athletic activities such as running, diving, bicycling and gymnastic vaulting. Although these behaviors differ greatly in character, our handcrafted systems all work in essentially the same way and are built from a common suite of components.

Our control systems use a state machine: an algorithm implemented in software that determines what each joint should be doing at every moment and then, like the conductor of an orchestra, ensures that the joints perform those functions at appropriate times. Running, for example, is a cyclic activity that alternates between a stance phase, when one leg is providing support, and a flight phase, when neither foot is on the ground. During the stance phase, the ankle, knee and hip of the leg that is in contact with the ground must provide support and balance. When that leg is in the air, however, the hip has a different function—that of swinging the limb forward in preparation for the next touchdown. The state machine selects among the various roles of the hip and chooses the right action for the current phase of the running motion.

Associated with each phase are control laws that compute the desired angles for each of the 30 joints of the simulated human body. The control laws are equations that represent how each body part should move to accomplish its intended function in each phase of the motion. To move the joints into the desired positions, the control system computes the appropriate torques with equations that act like springs, pulling the joints toward the desired angles. In essence, the equations are virtual muscles that move the various body parts into the right positions.

To simplify the problem of specifying the control laws, several limbs are often used in a synergistic fashion. For example, the ankle and knee joints of the simulated runner work together to push off the ground during the stance phase. Wherever possible, the control laws use the passive behaviors of the system to achieve a desired effect. Our assumption is that humans are efficient, and thus energy-conserving, passive behaviors will more closely mimic how people move. For example, during the stance phase, the runner's knee acts as a spring, both compressing to store energy and then extending to release that energy.

Simulated motion can also be made more natural by using limbs that are idle in a particular phase to reduce disturbances caused by the motion of other body parts. For the simulation of running, control laws make the athlete's arms swing in opposition to the legs to decrease yawing of the body.

Advantages and Disadvantages

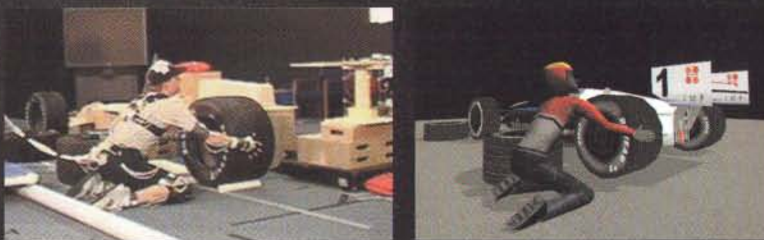
As a technique for synthesizing human motion, simulation has two potential advantages over keyframing and motion capture. First, simulations can easily be used to produce slightly different sequences while maintaining physical realism—for example, a person running at four meters per second instead of five. Merely speeding up or slowing down the playback of another type of animation can spoil the naturalness of the motion. Second, real-time simulations allow interactivity, an important feature for virtual environments and video games in which artificial characters must respond to the actions of an actual person. In contrast, applications based on keyframing and motion capture select and modify motions from a precomputed library of movements.

One drawback of simulation, however, is the expertise and time required to handcraft the appropriate control systems. My students and I have begun to address this issue by developing a library of software modules that can be combined to produce new movements. Specifically, we have constructed a set of four basic control systems for leaping, tumbling, landing and balancing. The computer can combine these systems to create more complex maneuvers, such as somersaults and dives.

We have also begun to explore ways to adapt existing behaviors to new characters. The process is difficult because control systems are tuned for the dynamic properties of a particular model. In general, a system developed for an adult does not work for a child. To modify an existing behavior for a new character with significantly different physical properties, we have developed optimization techniques for adjusting the control system. For example, we have taken our control system for a running man and successfully adapted it to a woman and to a four-year-old child by taking into account the different masses of the body parts of the three models and many other parameters.

In recent years, researchers have worked on simulation-based methods that generate motion without requiring the construction of a handcrafted control system. Several investigators have treated the synthesis of movement as a trajectory optimization problem. This formulation treats the equations of motion and significant features of the desired action as constraints and finds the motion that expends the least amount of energy while satisfying those restrictions. To simulate jumping, the constraints might state that the character should begin and end on the ground and be in the air in the middle of the mo-

Motion Capture



Much of the motion in commercial animation is generated by using data "captured" from the movements of a human subject. A computer can "replay" that recorded information to animate a synthetic character.

Simulation

In simulation the laws of physics are used to synthesize the movement of virtual people and objects.

COURTESY OF MICROSOFT MOTION CAPTURE GROUP (*left and center*); JAMES O'BRIEN AND VICTOR ZORDAN (*right*)

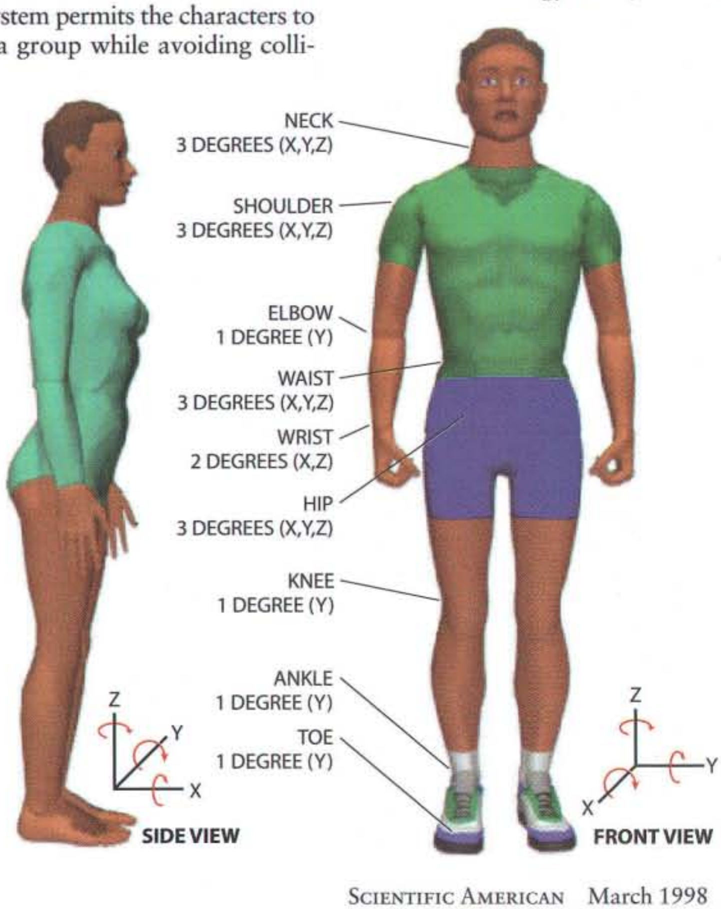
tion. The optimization software would then automatically determine that the character must bend its knees before jumping to get the maximum height for the minimum expenditure of energy. Another approach finds the best control system by automatically searching among all the possibilities. In the most general case, this technique must determine how a character could move from every possible state to every other state. Because this method solves a more general problem than that of finding a single optimum trajectory from a certain starting point to a particular goal, it has been most successful in simple simulations and for problems that have many solutions, thereby increasing the probability that the computer will find one. Fully automatic techniques are preferable to those requiring manual design, but researchers have not yet developed automatic methods that can generate behaviors for systems as complex as humans without significant prior knowledge of the movement.

Although control systems are difficult to construct, they are relatively easy to use. An animator can execute a simulation to generate motion without possessing a detailed understanding of the behavior or of the underlying equations. Simulation enables control over the general action but not the subtle details. For instance, the animator can dictate the path for a bicycle but cannot easily specify that the cyclist should be riding with a cheerful, lighthearted style. This limitation could be overcome in part by using simulation to generate the gross movements automatically and then relying on keyframing or on motion capture for the finer motions, such as facial expressions.

Simulated motion can be made more realistic through the addition of secondary, passive elements that move in response to the main character. A trampoline needs to deform as a gymnast bounces on it. Sand must compress to show footprints from a runner on the beach. Clothing—which can be modeled as a collection of points connected by springs—should move in reaction to the person wearing it.

Another feature of simulation is that it allows the animator to have high-level control over scenes involving crowds. A layer of software inserted above the control system permits the characters to move as a group while avoiding collisions with one another. This layer computes a desired velocity for each individual based on the proximity of other members in the group and obstacles in the environment. That velocity information is then used by the control system for locomotion.

My students and I chose to study athletic activities such as running and bicycling because the dynamics of these actions constrain the motion, thereby limiting the search for control laws. This property is most evident in our simulation of vaulting. The gymnast is airborne for much of the maneuver, and the control laws can influence the internal motion of the joints but not the angular momentum of the gymnast, which must



COMPUTER MODELS of the human body consist of rigid body parts connected by rotary joints that are assigned one, two or three degrees of freedom with respect to the x, y, z coordinate axes.

JESSICA K. HODGINS

Alan Alda Meets Alan Alda 2.0

Computers have had great difficulty in synthesizing the subtle facial expressions and movements that can convey megabytes of information about a person's mood, feelings and personality. But recent animations of "virtual" people that talk and display basic emotions do hint at the intriguing potential of the technology. Last fall, at the request of SCIENTIFIC AMERICAN *Frontiers,* Lamb & Company, a computer animation firm in Minneapolis, began building a digital "twin" of Alan Alda, the Emmy-award-winning actor. The synthetic Alda would meet its biological counterpart on television, speaking sentences the real Alda had never spoken.

For the ambitious project, Lamb built a computer model of Alda's head by using measurements obtained from laser scans of the real actor. The digital model consists of a "wire mesh" made up of roughly 12,000 tiny, contiguous polygons that represent the basic three-dimensional geometry of the actor's head. On top of this software sculpture lies a texture map containing information such as the coloration and smoothness of Alda's skin.

The laser scans were of four poses: a basic view of Alda showing no emotion and still shots of the actor grinning, grimacing and half-smiling. Using this information, Lamb created other expressions by putting together and finessing different combinations of the four poses. The Lamb animators also developed static views of the virtual Alda uttering different phonemes: basic units of speech such as the "k" sound in "cat." This work was essentially done by hand, although computer tools provided rudimentary assistance in modifying the software models. All told, the Lamb animators created a library of about 60 views, which could then be blended to create additional variations.

ALAN ALDA and his "virtual" twin will interact in an upcoming episode of SCIENTIFIC AMERICAN *Frontiers*. The digital Alda (*inset at right*) consists of a computer model of the actor's head, complete with synthesized voice. On location (*above*), a placeholder apparatus (*center*) was used. During the filming of the program, the real Alda conversed with the apparatus (*far right*), which will be replaced by the virtual Alda when the synthetic and real footage are composited.

To bring motion and "life" to the software model, Lamb worked with speech of the actor's voice that a laboratory in Japan had synthesized by separating and resequencing phonemes from a recording of the real Alda speaking. One painstaking task was to make the virtual Alda's lips move in synchronization with that simulated audio.

The work required the meticulous scrutiny of videotapes of the real Alda in action. "We watched how he moves his mouth as he makes an 'oh' or 'ooh' sound. We tried to get a feel for how his face should move," says Jim Russell, technical director on the project. Just as helpful were mirrors that the Lamb animators used to study the

obey the physical law for conservation of momentum. Running is a more complicated action because the synthetic athlete is in contact with the ground much of the time, and the joint torques computed by the control laws directly affect many of the details of the motion. Thus, much more effort went into tuning the motion of the runner than that of the gymnast. This observation has implications for the simulation of simple human movements in which style plays a large role, such as gesturing and fidgeting. When the physics involved do not constrain the gross characteristics of a motion, a computer can synthesize the movement successfully but in a way that appears unnatural. Such behaviors will require control laws that incorporate additional rules based on observations and measurements of people's movements.

Real Enough?

The criteria for assessing the quality of simulated human motion depend on the application. Virtual environments for training, collaboration and entertainment require motion with sufficient variety, expressiveness and realism to create a feeling of immersion for the user. Some virtual environments will be more compelling with realistic motions. For these applications, the grand challenge is a Turing test: Is simulated motion as natural as that of real people, at least when both types of movement are played back through the same graphical model? Preliminary experiments indicate that the answer to this question is, for the time being, no. But some viewers do choose simulated motion as more natural when both types are viewed with the human bodies removed so that the movement is shown only as dots located at the joints.

In contrast to applications such as virtual environments that are primarily visual, scientific and engineering problems require simulations that are validated by rigorous comparison with human data. One simple assessment is to compare real and synthesized video footage. Investigators also use biomechanical data for ground reaction forces, flight times, velocity and step length to

NANCY POLLARD

JOGGING MAN, WOMAN AND CHILD require modification of the simulation to take into account the dissimilar masses of the body parts of the three models and many other parameters.

VAULTING is a simpler action to simulate than running because the gymnast is airborne for much of the maneuver, during which her body must obey the physical law for conservation of angular momentum.

PHOTOGRAPHED BY SAM OGDEN AT THE MUSEUM OF FINE ARTS, BOSTON; "DIGITAL ALAN" BY LAMB & COMPANY

motion of their own faces when they spoke certain words.

The process of creating each frame in the synthesized footage might have been prohibitively time-consuming without the aid of keyframing, a technique in which an animator specifies just the key positions of an object. A computer then interpolates between those locations to fill in the remaining information. To synthesize a character saying the word "sell," for example, an animator might use still shots of the person saying "ess," "eh" and "ell," and the computer would then create the sequence of in-between frames automatically. Lamb also used this technique for other facial motions, such as when the virtual Alda blinks its eyes. Thanks to keyframing, Russell estimates that roughly only 20 percent of the work to generate the frames had to be done by hand.

But such automated techniques could not ensure that the virtual Alda would move naturally. Doing so required weeks of manual fine-tuning by seasoned animators. "Humans have so many irregularities of facial movement and speech," says Russell, who has more than 10 years of experience in computer animation. "The tweaking takes about 10 times the effort of getting the basic motion down."

The final footage, which consists of about 2,500 frames, took Lamb several months to synthesize. This arduous effort and the extensive computing required to bring a virtual Alan Alda to life for less than two minutes of television air time may make the salary of the real actor seem like a bargain in comparison. And, from all appearances, the digital Alda will need a great deal of additional refinement before contending for an Emmy of its own anytime soon.

—Alden M. Hayashi

Join both the virtual and real Alan Aldas in "The Art of Science," an upcoming episode of the SCIENTIFIC AMERICAN Frontiers *program. For time and channel, check your local listings.*

measure how closely the simulation resembles human motion.

As researchers improve their understanding of control systems, they will begin to answer important scientific and engineering questions with the technology. Physical therapists could benefit from insight into the gait irregularities that arise with particular injuries. Bicycle designers would like to determine the efficiency with which a rider can propel novel frame designs without having to construct prototypes. Diving coaches would like to know whether a particular person is strong enough to perform a new maneuver. Although the inherent complexity of control systems makes their design difficult, the strong scientific foundation of the technology makes it potentially suitable for these and other scientific applications. SA

The Author

JESSICA K. HODGINS is an assistant professor in the College of Computing at the Georgia Institute of Technology. Hodgins received her Ph.D. in computer science from Carnegie Mellon University in 1989 and was a postdoctoral fellow at the Massachusetts Institute of Technology Artificial Intelligence Laboratory and the IBM Thomas J. Watson Research Center. Her research focuses on the coordination and control of dynamic physical systems, both natural and man-made, and explores techniques that may someday allow robots and animated creatures to plan and control their actions in complex and unpredictable environments. She has received a National Science Foundation Young Investigator Award, a Packard fellowship and a Sloan fellowship.

Further Reading

SPACETIME CONSTRAINTS. Andrew Witkin and Michael Kass in *Proceedings of SIGGRAPH 88*. ACM, 1988.

SPACETIME CONSTRAINTS REVISITED. J. Thomas Ngo and Joe Marks in *Proceedings of SIGGRAPH 93*. ACM, 1993.

ANIMATING HUMAN ATHLETICS. Jessica K. Hodgins, Wayne L. Wooten, David C. Brogan and James F. O'Brien in *Proceedings of SIGGRAPH 95*. ACM, 1995.

DIGITAL CHARACTER ANIMATION. G. Maestri. New Riders Publishing, 1996.

ADAPTING SIMULATED BEHAVIORS FOR NEW CHARACTERS. Jessica K. Hodgins and Nancy S. Pollard in *Proceedings of SIGGRAPH 97*. ACM, 1997.

Examples of simulated human movement can be seen on the World Wide Web at http://www.cc.gatech.edu/gvu/animation/ The site contains work by the author and her students and postdoctoral researchers David Brogan, Wayne Wooten, James O'Brien, Deborah Carlson, Victor Zordan, Ronald Metoyer, Nancy Pollard and Robert Sumner.

WAYNE WOOTEN

The Caiman Trade

by Peter Brazaitis, Myrna E. Watanabe
and George Amato

A small, sleek handbag is prominently displayed in the showcase of a posh Madison Avenue store in New York City. Its $3,700 price tag tells the cognoscenti that the bag is made of real crocodilian skin, not printed calfskin. Only a handful of experts in the world may look at this handbag and know that although the glossy front is made of high-quality, legal American alligator skin, the sides are made from cheaply produced caiman leather. And the latter may have come from the largely contraband trade in skins.

Traffic-USA, a branch of the World Wildlife Fund that tracks the trade in contraband wildlife products, conservatively estimates the worldwide market for crocodilian skins at 1.5 to two million a year. Yet according to Don Ashley, a trade consultant in Tallahassee, Fla., in 1993 only about a million of those skins had legal documentation from the country of origin. So up to half of the skins that make up those expensive handbags, wallets and belts may have been harvested from wild animals, in violation of national or international laws. The bulk of these illegal skins comes from members of the genus *Caiman.*

The "alligator look" has historically projected luxury and affluence. Though always in style, the look is touted by the fashion industry in cycles of three to five years, periodically increasing the demand for all crocodilian-skin products. Figures for international trade in caiman skin, legal or illegal, were unavailable until 1993 (and these are still the only numbers available). Whereas "user" countries such as the U.S. reported the import of 648,847 skins, the World Conservation Union, known as the IUCN, found that only 556,132 skins had been legally exported that year. Clearly, some of those imported skins had fraudulent documents. The IUCN further estimated the actual trade in caiman hides to be at least one million a year, nearly 50 percent greater than the legal output.

The illegal trade in caimans is pervasive because it is

The contraband trade in caiman skins shows how "sustainable utilization" of endangered species fails to sustain them

FRANS LANTING *Minden Pictures*

YACARÉ CAIMAN, boasting an intricate pattern of scales on its sides, is a favorite target of hunters. The animals are easy to kill during the dry season, when they congregate in a few shallow pools.

so lucrative. Crocodilians may be ranched or farmed, as many in the U.S. are. The handbag in the Madison Avenue shop was made from the belly skin of an American alligator that may have netted the farmer or legal hunter several hundred dollars. The caiman skin—whether legally acquired from one of the ranches now mushrooming in Central and South America or illegally hunted from surrounding areas—may have cost anywhere from $5 to $50. But products made of caiman skin sell for as much as those of alligator skin.

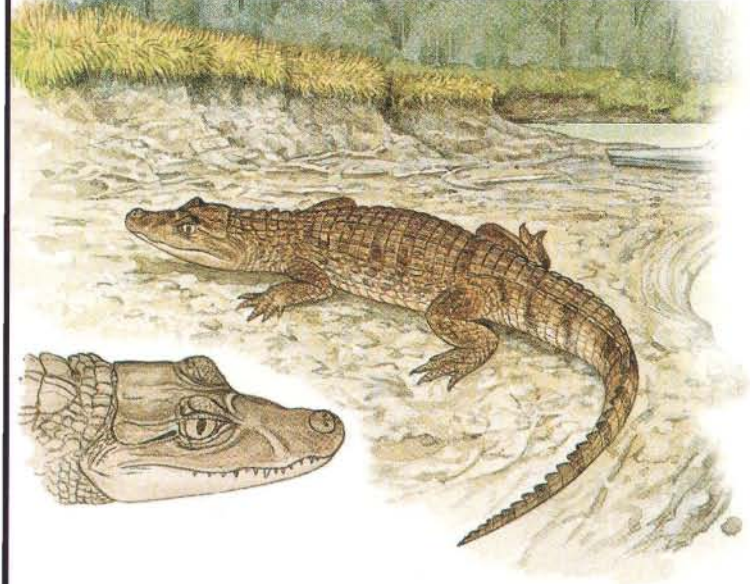
Classifying Caimans

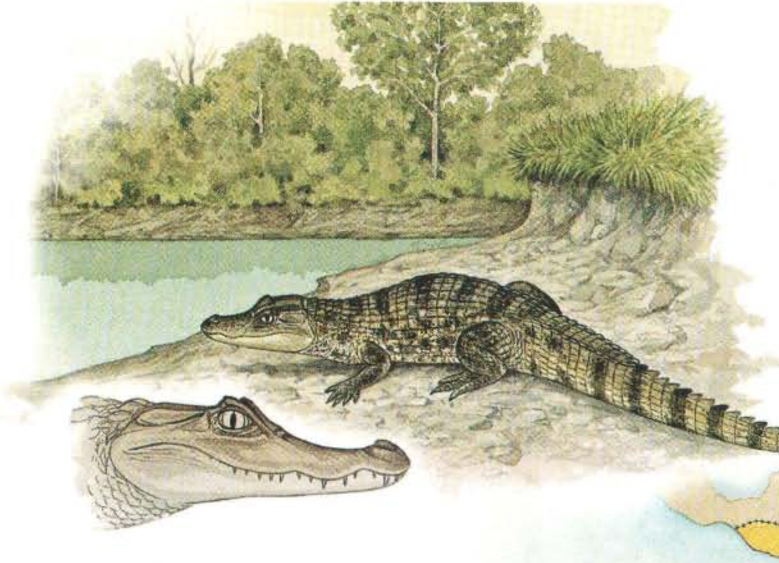
The crocodilians are members of the archosaurs, or ruling reptiles, the group that dominated the earth 200 million years ago. Most of the roughly 23 species of crocodilians are threatened or endangered, mainly as a result of habitat loss and excessive hunting.

Modern crocodilians are divided into three families: the true crocodiles (Crocodylidae), the gharials (Gavialidae) and the alligators (Alligatoridae). The last includes the American alligator, its near relative the diminutive Chinese alligator and the caiman. Caimans inhabit the freshwater rivers, streams, lakes, ponds and wetlands throughout tropical Central and South America and some of the Caribbean islands.

Three genera constitute the family of caimans. The first contains only one species, the giant—and now critically endangered—black caiman, *Melanosuchus niger.* The second, *Paleosuchus,* contains two species with such heavy bony plates within their scales that their hides are commercially unusable. The third, *Caiman,* contains a variety of species, most of them fairly small animals ranging from 1.2 to 2.8 meters (four to nine feet) in length.

Much controversy rages over the classification of caimans. Until a recent survey conducted by one of us (Brazaitis) and his colleagues, no studies correlated the dis-





ROBERTO OSTI

DUSKY CAIMANS grow to two meters (6.5 feet) in length and occupy a spectrum of habitats. Until quite recently, they were safe from the pressures of commerce because their hides are of relatively poor quality. But as species with more desirable skins become rare, the demand for dusky caiman skin is increasing. Wild populations in Colombia, where most hides come from, have fallen.

COMMON CAIMANS are wily and opportunistic, favoring varied habitats and diets. But a recent survey in Brazil found no common caimans at all in some regions where they had once been, in fact, common. The animals grow up to 2.5 meters (eight feet) in length. Most legal skins of common caimans come from Venezuela, and most illegal skins from Brazil. Some skins show high levels of lead. The habitat is also contaminated with mercury.

tribution of caimans in South America's river systems with their scale patterns, morphology and DNA characteristics. Robert Madden of Marymount College in Tarrytown, N.Y., has statistically analyzed the new data on structure, scale and color patterns; in a complementary study, another of us (Amato) and his colleagues at the Wildlife Conservation Society and the American Museum of Natural History have studied DNA from caiman tissues and blood. Contrary to the prevailing wisdom that the genus *Caiman* held only two distinct species, their results support the existence of four. Furthermore, the researchers supplied specifications for distinguishing among the species, a great aid to law-enforcement officials who were confounded by the variety within the species.

The common caiman (also called the spectacled caiman), *Caiman crocodilus,* mainly occurs in the drainage basins of the Amazon and Orinoco rivers. The dusky or brown caiman, *Caiman fuscus,* ranges from Mexico down to Peru, being isolated from the common caiman by the Andes Mountains. Both species are separated by scattered uplands and the Brazilian shield from the Yacaré caiman, *Caiman yacare*. The fourth member of the group, the broad-snouted caiman, or *Caiman latirostris*, roams river systems along the eastern uplands, coastal and southern tropical regions of South America.

A Watchful Parent

A hundred years ago explorers traveling through Brazil spoke of caimans stretched out one next to another, carpeting the banks of the Amazon River. Black caimans, reaching almost seven meters in length, were especially numerous, and occasionally picked off humans who ventured too close to the water's edge. Today caimans are abundant in some isolated regions but extirpated (that is, locally extinct) in others.

The annual wet season typically signals the time for breeding. Males vocalize and establish territories, within which one or more females build their nests. Sometimes more than two meters wide and half a meter high, the nests lie on high ground above flood levels but not far from the water. They are made of grasses, twigs and mud, usually in dense thickets or brush.

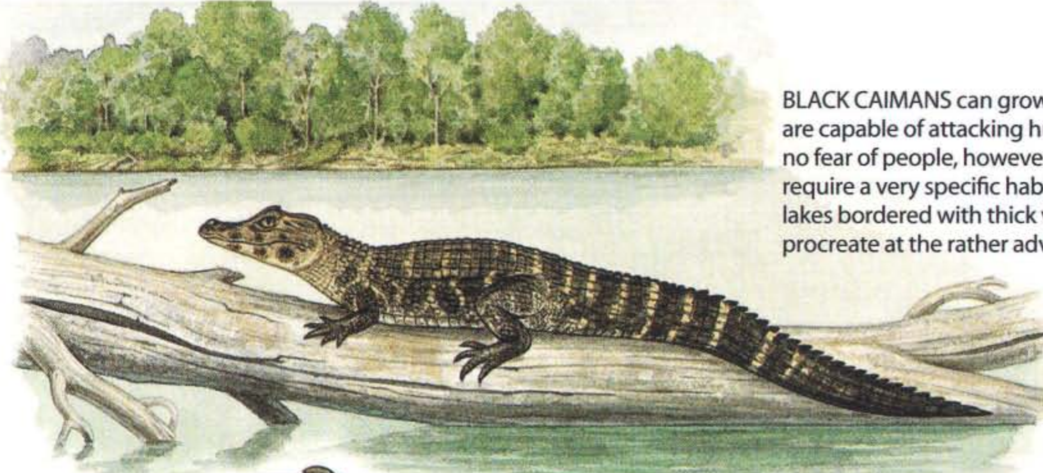
The decomposing organic material of the nest provides temperatures of 28 to 32 degrees Celsius (82 to 90 degrees Fahrenheit) throughout the incubation period of about eight to 10 weeks. Females lay 20 to 30 eggs in the nest at the peak of the rainy season, guarding them vigilantly, while males often patrol the vicinity. When hatching time nears and the baby caimans call from within the eggs, a female will excavate her nest to liberate the hatchlings, sometimes carrying them gently in her mouth to the seclusion of a nearby nursery pool.

Baby caimans stay with their parents for protection throughout most of the first year and are often seen riding on their backs. Juveniles and young adults feed on terrestrial invertebrates and, as they grow, large numbers of snails (a common food for many species of young crocodilians worldwide). Snails harbor the *Schistosoma* parasite, and some anecdotal evidence suggests that caiman predation on snails may help control schistosomiasis, a major tropical disease in humans and domestic animals.

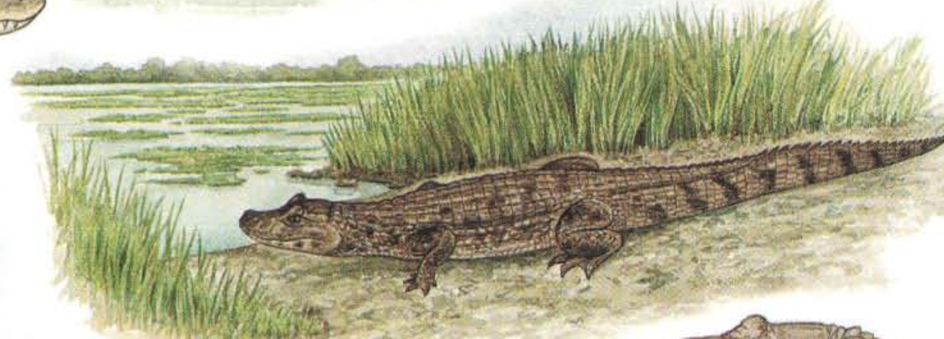
Despite the parental attention, the hatchlings often become a meal for many other species. Turtles, predatory fish such as piranhas, aquatic birds and snakes hunt very young caimans. If statistics from wild American alligator populations are taken as examples, perhaps 50 percent or more of all hatchlings are killed by predators during the first year of life.

Those caimans that grow larger eat fish that are not likely to find their ways into the human diet, such as armored catfish and some eels, as well as potentially dangerous species, such as piranhas. Without population control by these predators, many of these fish species would outcompete those eaten by humans. Moreover, caimans readily switch to feeding on whatever is available. Thus, they are well adapted to an environment in which water resources vary tremendously over the year.

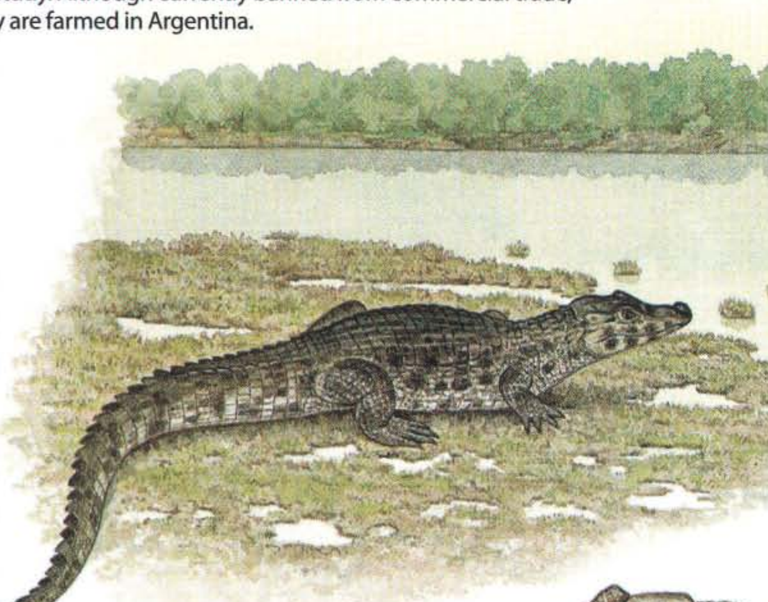
In the rainy season, caimans fan out into swamps and forests; in the dry season, they excavate the few pools that re-



BLACK CAIMANS can grow to six meters (19 feet) and are capable of attacking humans. The animals have no fear of people, however, and are easy to kill. They require a very specific habitat: open rivers or oxbow lakes bordered with thick vegetation. They begin to procreate at the rather advanced age of 12 years. As a result, once they are extirpated, populations do not recover.



BROAD-SNOUTED CAIMANS are nearly extinct because of habitat loss and pollution, and also because their hides provide a soft leather with a "marshmallow" feel. The animals may grow to 2.3 meters (7.5 feet). Little is known about these creatures in the wild because so few have been located for study. Although currently banned from commercial trade, they are farmed in Argentina.



main and congregate there. By doing so, not only do they make themselves a home, they also provide a refuge for fish, amphibians, aquatic invertebrates and plants. Often the only sources of aquatic food, the ponds attract migratory birds that feast on the fish and invertebrates.

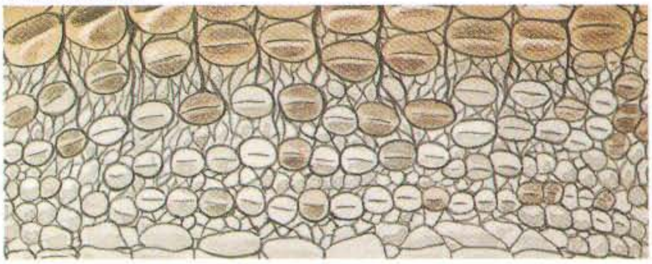
Caimans may live to be 65, although very few actually do. Adult caimans are sometimes eaten by anacondas. But human predation is a far more important factor in the species' demise. Local consumers take the medium-to-large animals for food; hunters kill larger animals for their skins.

An IUCN report notes that wild populations of the common caiman are declining in half of the 16 countries where it occurs. The Yacaré, broad-snouted and black caimans are depleted, the latter two severely, in all their range countries. This situation is a result of human encroachment on their habitat and nearly half a century of poaching.

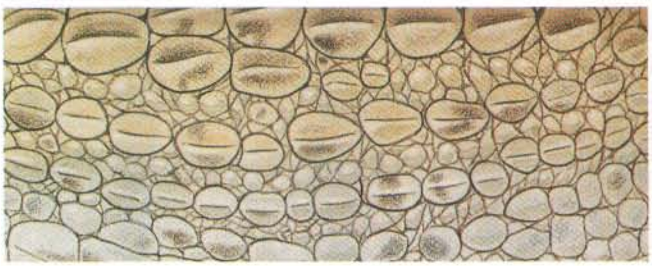
The international trade in all wildlife, including caimans (whether live animals for pets or dead ones for meat, skins and other products), is regulated by CITES, the Convention on International Trade in Endangered Species of Wild Fauna

YACARÉ CAIMANS reach a maximum of 2.5 meters (eight feet) in length. They are the primary source of skin for the crocodilian leather industry. Although banned from trade in the U.S., their import may soon be legalized. The species is widespread, but wild populations are declining.

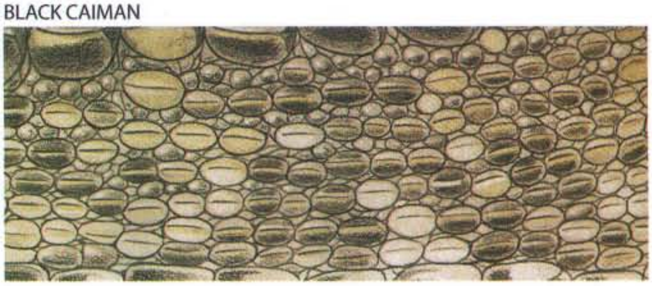
DUSKY CAIMAN



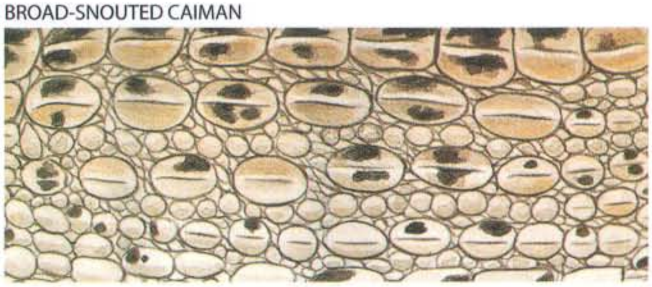
COMMON CAIMAN



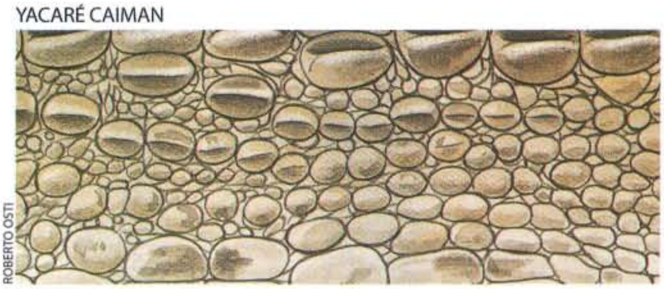
BLACK CAIMAN



BROAD-SNOUTED CAIMAN



YACARÉ CAIMAN



ROBERTO OSTI

UNIQUE PATTERNS of scales on the flanks are useful for identifying caiman skins. Dusky caimans have uniform rows of oval scales alternating with rows of small, beadlike scales. Common caimans have narrow flanks with strongly keeled (that is, raised), large scales laid in rows, separated by soft, creased skin and tiny scales. Black caimans have wide flanks with numerous rows of slightly elongated, poorly keeled scales. The softest leather from these animals is that of the smooth, white belly (*not shown*). Broad-snouted caimans have a few rows of large, bony scales alternating with small, bony scales. Yacaré caimans have wide flanks with mostly smooth, round scales that lie in close rows separated by a chainlike pattern of creased skin. For all the skins depicted, the animals' heads point to the left.

and Flora. National laws, such as those in the country of origin of the animals or the Endangered Species Act in the U.S., also apply. CITES uses the available data on a species to classify it into one of three groups. Appendix I animals are not allowed in international commerce. Appendix II animals are permitted only with CITES documentation issued by government authorities in the country of origin (or, if the animal originates in a different country, with appropriate permits for reexport). Appendix III animals may be traded without such stringent requirements. The broad-snouted caiman, the Rio Apaporis caiman (a subspecies of the common caiman) and the black caiman are listed as CITES Appendix I species. The remainder of the caimans are CITES Appendix II.

Differentiating what is legal from what is not gets very messy. Many countries export skins that could never have originated there—but with legal documentation claiming otherwise. It may transpire, for instance, that a skin exported from Colombia is declared as a common caiman when it is actually a Yacaré, which does not occur in Colombia. There may even be a brisk trade in illegal documentation: CITES export forms are sometimes reported lost.

When the U.S. Fish and Wildlife Service seizes caiman skins or products on suspicion that they are illegal, it often engages the services of a caiman specialist. Brazaitis is one such taxonomist in the U.S. To prosecute the owners, the service must identify the species of each item in the shipment—and there may be thousands. That is not an easy task, because the Yacaré and the various subspecies of the common caiman have skins that look extraordinarily alike. Once tanned and dyed, they no longer maintain the color patterns unique to the species. To make matters more difficult, odd strips of skin pieced together on leather goods may not include definitive characteristics for identification.

In 1969 Brazaitis and F. Wayne King of the University of Florida at Gainesville, both then at the Bronx Zoo, began to develop tools to differentiate between the species. By 1984 they had part of the problem licked. Caimans have thick, bony plates called osteoderms within the belly scales, which clearly distinguish the group from other crocodilians. The flanks, between the front and hind legs, along with the sides of the tails, are less bony and make supple leather. These pieces, used most often in products, are key to distinguishing endangered caimans. But more information was needed.

In the mid-1980s the United Nations Environmental Program, along with individual governments and CITES, funded an ambitious project to identify and quantify caiman populations in Brazil, Bolivia and Paraguay. Brazaitis led a Brazilian team that included, among many biologists, George Rebêlo of the National Institute for Investigation of the Amazon (INPA) and Carlos Yamashita of the Brazilian Institute for Environment and Renewable and Natural Resources (IBAMA). The team located caiman populations and collected blood, muscle, skin and other tissues from animals in the wild, as well as some destined for local residents' stew pots and ones found dying of pollution or of injuries inflicted by hunters. (The wild animals were captured and released after study. If they were extremely sick, we humanely euthanatized them.)

The tissue samples are useful not only for classifying caimans but also for providing an accurate means of identifying untanned skins. In the near future, the U.S. Fish and Wildlife Service will probably be able to utilize DNA evidence such as that obtained by Amato to identify tanned leather from endangered species.

Over the course of five years, our team found that, much to our surprise, some previously known populations of caimans were entirely gone. Of the remainder, the bulk of animals at the sites studied were very young. Few adult caimans were seen. Because we surveyed the animals in both the rainy season, when caimans are dispersed in the abundant waterways, as well as in the dry season, when the animals gather in great densities at the few pools available, we began to realize that there were very few mature adult caimans. The lack of adults means that few animals of reproductive age are left to continue the species; moreover, without parental protection, few of the juveniles will survive. It also implies a serious threat to the region's ecology. The waterholes excavated by caimans are often the only source of food available not only for the caimans but also for passing birds and mammals.

DON ASHBY Sygma

FASHION INDUSTRY promotes crocodilian skins in cycles of about four years, increasing the pressures on wild populations of caimans.

Furthermore, Elizabeth Odierna, then a student at Manhattan College in the Bronx, found that some of the caiman tissues were contaminated with dangerous levels of lead. This discovery came as a complete surprise: her true goal had been to check for mercury, used in many areas of caiman habitats to extract gold ore. Brazilian biologists had earlier discovered high levels of mercury in virtually all species of fish eaten by caimans and humans throughout the Amazon; our field tests confirmed the contamination of caiman habitats by mercury. Although the effect of mercury and lead on caimans is not known, one can only presume that the metals are as detrimental to these creatures as to humans.

Unsustainable Utilization

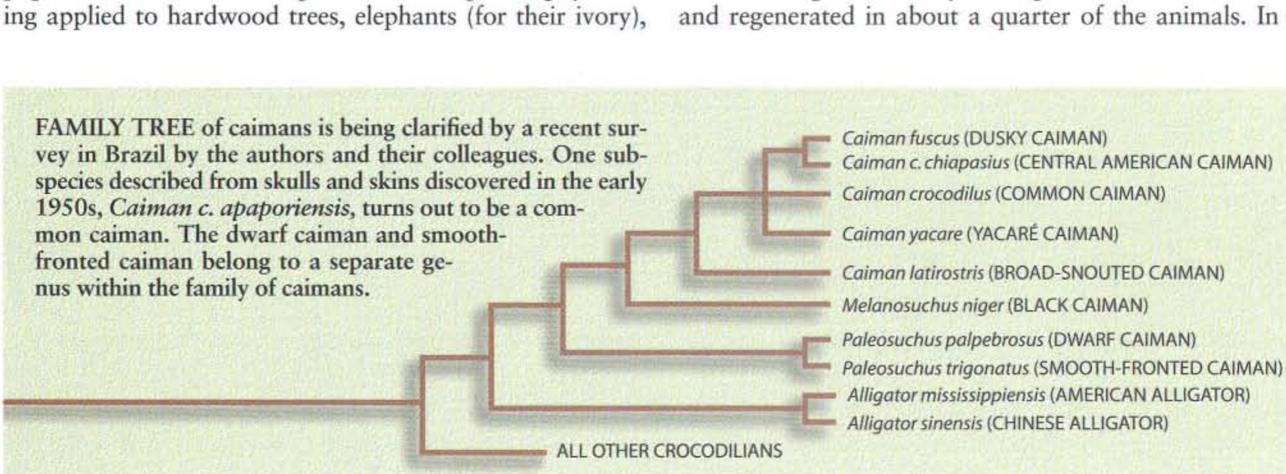
The watchwords for many crocodilian biologists today are "sustainable utilization," defined as use of the species so that it maintains an economic value to the local human population and is therefore preserved. This philosophy is being applied to hardwood trees, elephants (for their ivory), whales (for meat)—to just about every endangered or threatened species. Almost the entire conservation effort for crocodilians is directed toward using them for luxury leather products and meat.

Currently investors and ranchers in South America are spearheading a movement to build caiman farms and ranches to allow legal exportation of skins. Farming involves taking live adults from the wild and keeping and breeding them in captivity. Some of the offspring are used for leather and meat, and the rest are raised into breeding adults. If properly run, a farm should be a closed system, with no importation of animals beyond the initial batch. In ranching, eggs are collected from wild nests and incubated, or hatchlings are taken from the wild. The young are raised until the skins would be about a meter in length, when they are harvested. The resource for a ranch is the wild; hence, the argument goes, ranching creates an incentive to preserve caiman habitats.

Reality is tragically distant from theory. According to trade consultant Ashley, Brazil already has 75 registered ranches. Whereas some farms and ranches are intensively managed and have pens and buildings housing the caimans, in some places a "farm" or "ranch" may be a wild area that is fenced in by a private owner. The caimans unfortunate enough to live there then become available for commerce. Sometimes adults are collected from already depleted wild populations to form the basis of a commercial operation. Where wild animals are accessible just across the fence, an improperly run farm or ranch can become a laundromat for illegally captured wild animals.

There is circumstantial evidence that some farmed animals are actually wild ones. Brazaitis recently examined shipments of caiman skin with documentation indicating their source to be Colombian farms. The skins, up to 1.6 meters in length—and obviously originating from much longer animals—had an interesting abnormality. The tips of the tails had been lost and regenerated in about a quarter of the animals. In the

FAMILY TREE of caimans is being clarified by a recent survey in Brazil by the authors and their colleagues. One subspecies described from skulls and skins discovered in the early 1950s, *Caiman c. apaporiensis*, turns out to be a common caiman. The dwarf caiman and smooth-fronted caiman belong to a separate genus within the family of caimans.



BRYAN CHRISTIE

wild, as many as 50 percent of young caimans may lose the ends of their tails to predators. But this rarely occurs on a farm, whose objective is to produce unblemished skins. Furthermore, farmed or ranched animals are usually harvested well before they reach even 1.5 meters in length.

Many conservationists see commercial farming and ranching of crocodilians as providing cash that will somehow be redirected into conservation. But they seem to believe countries with poor histories of enforcing wildlife laws will become completely law-abiding overnight. Investors are expected to provide food, medicine and care to a hatchling for about three years until it reaches market size and then to make a profit by selling it for only $50 or less. Ted Joanen, formerly with the Rockefeller Wildlife Refuge in Louisiana, points out that even in South America's impoverished economy such expectations are unrealistic. Moreover, the bulk of the revenue from the sale of skins does not make its way into enforcing wildlife laws or preserving habitat and wild populations. Rather it goes into investment profits, operating budgets or general government funds.

Once a prohibited species is put into trade, it instantly generates a new market. Without sufficient international controls and the ability to separate what is legal from what is not, the legalization serves simply to sanitize what remains largely an illegal trade. The critically endangered black caiman, ranched in Ecuador, is most likely soon to be allowed in international trade. But controlling the movement of illegal black caiman skins from cheap wild sources to products is simply impossible. The animals have relatively slow reproductive cycles, are easy to kill and require very specific habitats; once extirpated, they rarely return. As a result, even Ashley is concerned that when the trade begins, wild populations of the species will need to be closely monitored. Yet no program for monitoring wild populations is in place.

When a country decides to invest in sustainable utilization, the wild animals become tagged as resources. Typically all the available funding goes toward technology for farming, often with no effort being made to preserve the habitat. The animals of commercial value are selectively preserved as cash cows—not as biological functionaries in an ecosystem. Nowhere is this better seen than in China, which started farming the Chinese alligator in 1979. Although the farming operation is now successful, the wild alligator is vanishing: fewer than 1,000 animals remain in the wild. Nor can the wild population be regenerated from the farmed one; there is no natural habitat left to which to return the animals.

In principle, a farm can serve to breed a critically endangered species that is then returned to the wild. This has occurred in India, where saltwater and mugger crocodiles as well as gharials have been restocked by the Indian government and the Madras Crocodile Bank. It should be noted, however, that India legally protects wild populations and wildlife refuges and does not allow trade in crocodilians, so there is less incentive for the wild animals to be killed.

A Question of Science

Scientific research into crocodilian biology is heavily supported by the farming, ranching and leather industries. Nearly all the major funders of the Crocodile Specialist Group that advises IUCN—which in turn makes recommendations to CITES and individual governments—are members of the trade. With so many scientists dependent on the industry's support, species conservation is in danger of getting short shrift when pitted against sustainable utilization.

Industry consultants hold that increasing the trade of the species will increase the incentive to protect the species. Based on such reasoning, CITES has assigned an export quota of 600,000 caiman skins a year to Colombia, a nation that actually produces up to 450,000 (mainly *C. fuscus*) a year. The rationale is that Colombia will soon be able to increase its caiman exports by means of farming and ranching.

In 1973, when the Endangered Species Act was signed, the U.S. Fish and Wildlife Service classified the Yacaré caiman as endangered and forbidden in U.S. trade. Significant amounts of data on the species' abundance were not available, and as a result the crocodilian leather industry has made numerous efforts since the mid-1980s to have the Yacaré declassified. When its initial efforts failed, the leather industry sponsored a study that divided Yacaré into three subspecies based on tannery skins of unknown origins. It then argued that these so-called subspecies should be traded, an argument to which the U.S. Fish and Wildlife Service fortunately did not subscribe. (But CITES manuals currently list all three Yacaré subspecies, without offering effective means for law-enforcement officials to distinguish among them.) Data from our survey prove that these multiple races of Yacaré are entirely without foundation. Efforts to downlist the Yacaré continue, even though their wild populations (and those of all other species of caimans) are reported to be declining in all their ranges.

As pressure to open trade in all crocodilian species increases, there continues to be no way to tell whether a piece of skin was from a legally farmed or ranched animal or from a "protected" wild creature that was hit on the head while basking on the edge of its pool. Aside from animals kept on commercial farms and ranches, South and Central America's wild caimans may well be on their way to becoming a true relic—the side of a purse displayed in a museum. SA

The Authors

PETER BRAZAITIS, MYRNA E. WATANABE and GEORGE AMATO bring diverse skills to the study of caimans. Brazaitis is a forensic specialist in herpetology and has served with the Wildlife Conservation Society in New York City for 43 years. He is co-coordinator of the Crocodilian Advisory Group of the American Zoo and Aquarium Association. Watanabe, who is married to Brazaitis, is an animal behaviorist. She has studied American and Chinese alligators in their natural habitat and writes on biotechnology. Amato is director of genetics research and senior conservation biologist at the Wildlife Conservation Society. His major research projects are long-term studies of population differentiation, gene flow and evolutionary history of crocodilians, rhinoceri, Asian bovids and cervids, and other vertebrates.

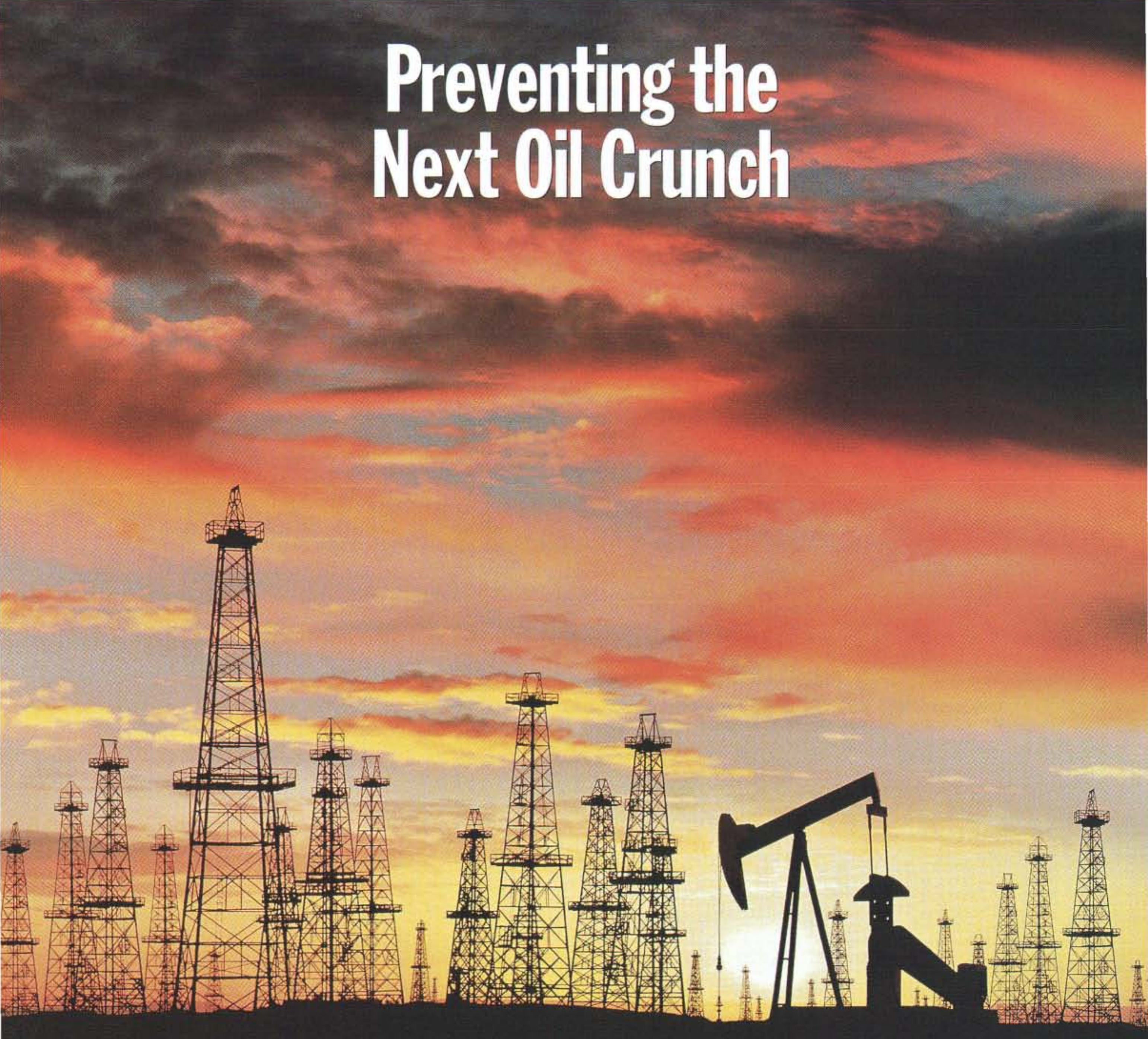
Further Reading

REPTILE LEATHER TRADE: THE FORENSIC SCIENCE EXAMINER'S ROLE IN LITIGATION AND WILDLIFE LAW ENFORCEMENT. Peter Brazaitis in *Journal of Forensic Science,* Vol. 31, pages 621–629; 1986.

CROCODILES AND ALLIGATORS. Edited by Charles A. Ross. Facts on File, 1989.

THREATS TO BRAZILIAN CROCODILIAN POPULATIONS. Peter Brazaitis, George H. Rebêlo, Carlos Yamashita, Elizabeth A. Odierna and Myrna E. Watanabe in *Oryx,* Vol. 30, No. 4, pages 275–284; 1996.

Preventing the Next Oil Crunch



BILL ROSS *Tony Stone Images*

Enough oil remains in the earth to fill the reservoir behind Hoover Dam four times over—and that's just counting the fraction of buried crude that is relatively easy to recover and refine. Little wonder, then, that as the world's economies have hit the accelerator in the past decade, the production of oil that powers them has also soared, reaching a record of 65 million barrels a day last year. The ample supply has kept oil cheap and has helped to lay inflation low.

But could this be the peak before the fall? The authors of the first article in this special report conclude that before the next decade is over the flood of conventional oil will crest, and production will enter a permanent decline.

These analysts marshal an impressive body of statistics to support their projections. If they are right, the world will need to move quickly to avoid the price hikes, recessions and political struggles that oil shortages—or threats of them—have historically provoked. But as explained here, there are alternatives. The industry can eke more out of aging oil fields and can sink new wells beneath deeper ocean waters. And the technology already exists to convert natural gas and oil sands, both still plentiful, into liquid fuels that are as cheap as those refined from oil. The means for an orderly transition away from crude oil appear to be nearly ready; all that is needed is the will, the time and the money. *—The Editors*

The End of Cheap Oil

Global production of conventional oil will begin to decline sooner than most people think, probably within 10 years

by Colin J. Campbell and Jean H. Laherrère

In 1973 and 1979 a pair of sudden price increases rudely awakened the industrial world to its dependence on cheap crude oil. Prices first tripled in response to an Arab embargo and then nearly doubled again when Iran dethroned its Shah, sending the major economies sputtering into recession. Many analysts warned that these crises proved that the world would soon run out of oil. Yet they were wrong.

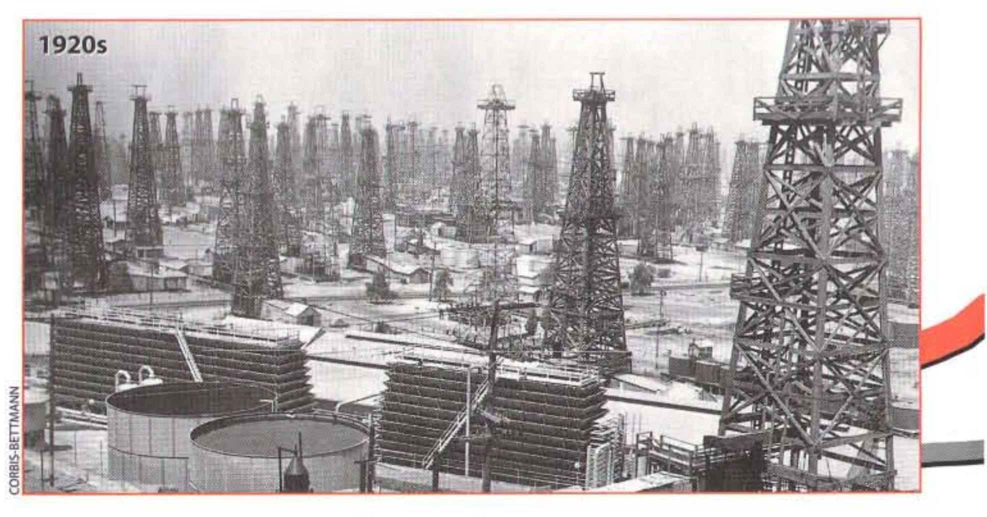
Their dire predictions were emotional and political reactions; even at the time, oil experts knew that they had no scientific basis. Just a few years earlier oil explorers had discovered enormous new oil provinces on the north slope of Alaska and below the North Sea off the coast of Europe. By 1973 the world had consumed, according to many experts' best estimates, only about one eighth of its endowment of readily accessible crude oil (so-called conventional oil). The five Middle Eastern members of the Organization of Petroleum Exporting Countries (OPEC) were able to hike prices not because oil was growing scarce but because they had managed to corner 36 percent of the market. Later, when demand sagged, and the flow of fresh Alaskan and North Sea oil weakened OPEC's economic stranglehold, prices collapsed.

The next oil crunch will not be so temporary. Our analysis of the discovery and production of oil fields around the world suggests that within the next decade, the supply of conventional oil will be unable to keep up with demand. This conclusion contradicts the picture one gets from oil industry reports, which boasted of 1,020 billion barrels of oil (Gbo) in "proved" reserves at the start of 1998. Dividing that figure by the current production rate of about 23.6 Gbo a year might suggest that crude oil could remain plentiful and cheap for 43 more years—probably longer, because official charts show reserves growing.

Unfortunately, this appraisal makes three critical errors. First, it relies on distorted estimates of reserves. A second mistake is to pretend that production will remain constant. Third and most important, conventional wisdom erroneously assumes that the last bucket of oil can be pumped from the ground just as quickly as the barrels of oil gushing from wells today. In fact, the rate at which any well—or any country—can produce oil always rises to a maximum and then, when about half the oil is gone, begins falling gradually back to zero.

From an economic perspective, when the world runs completely out of oil is thus not directly relevant: what matters is when production begins to taper off. Beyond that point, prices will rise unless demand declines commensurately.

HISTORY OF OIL PRODUCTION, from the first commercial American well in Titusville, Pa. (*left*), to derricks bristling above the Los Angeles basin (*below*), began with steady growth in the U.S. (*red line*). But domestic production began to decline after 1970, and restrictions in the flow of Middle Eastern oil in 1973 and 1979 led to inflation and shortages (*near* and *center right*). More recently, the Persian Gulf War, with its burning oil fields (*far right*), reminded the industrial world of its dependence on Middle Eastern oil production (*gray line*).



Using several different techniques to estimate the current reserves of conventional oil and the amount still left to be discovered, we conclude that the decline will begin before 2010.

Digging for the True Numbers

We have spent most of our careers exploring for oil, studying reserve figures and estimating the amount of oil left to discover, first while employed at major oil companies and later as independent consultants. Over the years, we have come to appreciate that the relevant statistics are far more complicated than they first appear.

Consider, for example, three vital numbers needed to project future oil production. The first is the tally of how much oil has been extracted to date, a figure known as cumulative production. The second is an estimate of reserves, the amount that companies can pump out of known oil fields before having to abandon them. Finally, one must have an educated guess at the quantity of conventional oil that remains to be discovered and exploited. Together they add up to ultimate recovery, the total number of barrels that will have been extracted when production ceases many decades from now.

The obvious way to gather these numbers is to look them up in any of several publications. That approach works well enough for cumulative production statistics because companies meter the oil as it flows from their wells. The record of production is not perfect (for example, the two billion barrels of Kuwaiti oil wastefully burned by Iraq in 1991 is usually not included in official statistics), but errors are relatively easy to spot and rectify. Most experts agree that the industry had removed just over 800 Gbo from the earth at the end of 1997.

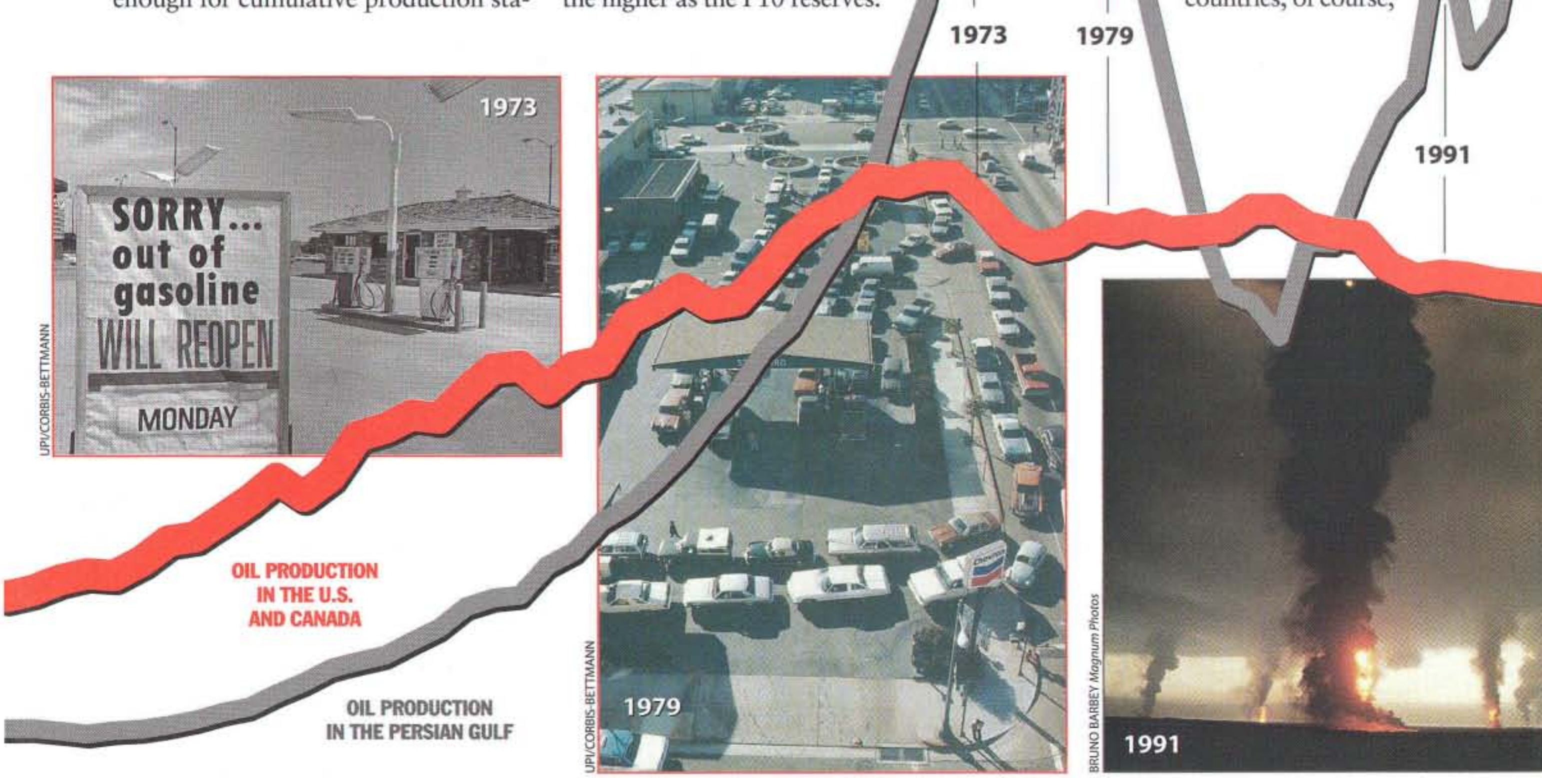
Getting good estimates of reserves is much harder, however. Almost all the publicly available statistics are taken from surveys conducted by the *Oil and Gas Journal* and *World Oil*. Each year these two trade journals query oil firms and governments around the world. They then publish whatever production and reserve numbers they receive but are not able to verify them.

The results, which are often accepted uncritically, contain systematic errors. For one, many of the reported figures are unrealistic. Estimating reserves is an inexact science to begin with, so petroleum engineers assign a probability to their assessments. For example, if, as geologists estimate, there is a 90 percent chance that the Oseberg field in Norway contains 700 million barrels of recoverable oil but only a 10 percent chance that it will yield 2,500 million more barrels, then the lower figure should be cited as the so-called P90 estimate (P90 for "probability 90 percent") and the higher as the P10 reserves.

In practice, companies and countries are often deliberately vague about the likelihood of the reserves they report, preferring instead to publicize whichever figure, within a P10 to P90 range, best suits them. Exaggerated estimates can, for instance, raise the price of an oil company's stock.

The members of OPEC have faced an even greater temptation to inflate their reports because the higher their reserves, the more oil they are allowed to export. National companies, which have exclusive oil rights in the main OPEC countries, need not (and do not) release detailed statistics on each field that could be used to verify the country's total reserves. There is thus good reason to suspect that when, during the late 1980s, six of the 11 OPEC nations increased their reserve figures by colossal amounts, ranging from 42 to 197 percent, they did so only to boost their export quotas.

Previous OPEC estimates, inherited from private companies before governments took them over, had probably been conservative, P90 numbers. So some upward revision was warranted. But no major new discoveries or technological breakthroughs justified the addition of a staggering 287 Gbo. That increase is more than all the oil ever discovered in the U.S.—plus 40 percent. Non-OPEC countries, of course,



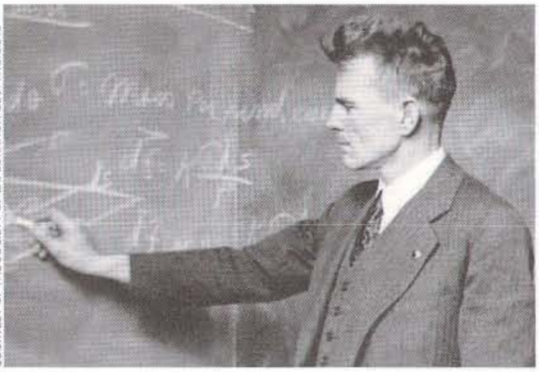
JENNIFER C. CHRISTIANSEN; SOURCE: JEAN H. LAHERRÈRE

UPI/CORBIS-BETTMANN

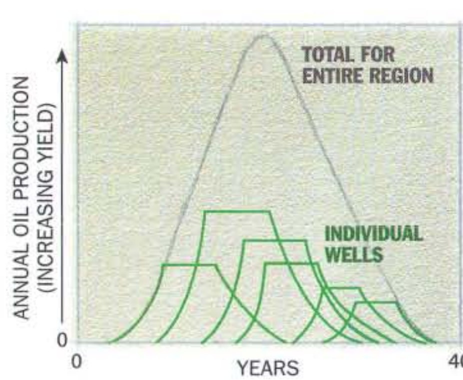
UPI/CORBIS-BETTMANN

BRUNO BARBEY *Magnum Photos*

COURTESY OF THE SOCIETY OF EXPLORATION GEOPHYSICISTS



FLOW OF OIL starts to fall from any large region when about half the crude is gone. Adding the output of fields of various sizes and ages (*green curves at right*) usually yields a bell-shaped production curve for the region as a whole. M. King Hubbert (*left*), a geologist with Shell Oil, exploited this fact in 1956 to predict correctly that oil from the lower 48 American states would peak around 1969.



JENNIFER C. CHRISTIANSEN; LAURIE GRACE

are not above fudging their numbers either: 59 nations stated in 1997 that their reserves were unchanged from 1996. Because reserves naturally drop as old fields are drained and jump when new fields are discovered, perfectly stable numbers year after year are implausible.

Unproved Reserves

Another source of systematic error in the commonly accepted statistics is that the definition of reserves varies widely from region to region. In the U.S., the Securities and Exchange Commission allows companies to call reserves "proved" only if the oil lies near a producing well and there is "reasonable certainty" that it can be recovered profitably at current oil prices, using existing technology. So a proved reserve estimate in the U.S. is roughly equal to a P90 estimate.

Regulators in most other countries do not enforce particular oil-reserve definitions. For many years, the former Soviet countries have routinely released wildly optimistic figures—essentially P10 reserves. Yet analysts have often misinterpreted these as estimates of "proved" reserves. *World Oil* reckoned reserves in the former Soviet Union amounted to 190 Gbo in 1996, whereas the *Oil and Gas Journal* put the number at 57 Gbo. This large discrepancy shows just how elastic these numbers can be.

Using only P90 estimates is not the answer, because adding what is 90 percent likely for each field, as is done in the U.S., does not in fact yield what is 90 percent likely for a country or the entire planet. On the contrary, summing many P90 reserve estimates always understates the amount of proved oil in a region. The only correct way to total up reserve numbers is to add the mean, or average, estimates of oil in each field. In practice, the median estimate, often called "proved and probable," or P50 reserves, is more widely used and is good enough. The P50 value is the number of barrels of oil that are as likely as not to come out of a well during its lifetime, assuming prices remain within a limited range. Errors in P50 estimates tend to cancel one another out.

We were able to work around many of the problems plaguing estimates of conventional reserves by using a large body of statistics maintained by Petroconsultants in Geneva. This information, assembled over 40 years from myriad sources, covers some 18,000 oil fields worldwide. It, too, contains some dubious reports, but we did our best to correct these sporadic errors.

According to our calculations, the world had at the end of 1996 approximately 850 Gbo of conventional oil in P50 reserves—substantially less than the 1,019 Gbo reported in the *Oil and Gas Journal* and the 1,160 Gbo estimated by *World Oil*. The difference is actually greater than it appears because our value represents the amount most likely to come out of known oil fields, whereas the larger number is supposedly a cautious estimate of proved reserves.

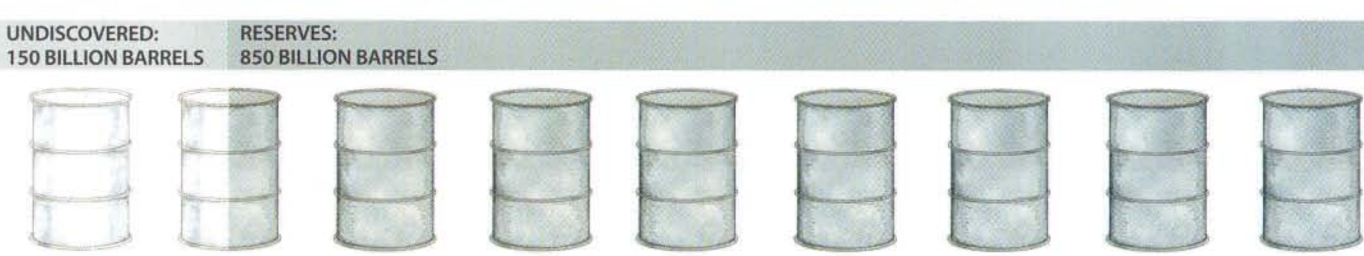
For the purposes of calculating when oil production will crest, even more critical than the size of the world's reserves is the size of ultimate recovery—all the cheap oil there is to be had. In order to estimate that, we need to know whether, and how fast, reserves are moving up or down. It is here that the official statistics become dangerously misleading.

Diminishing Returns

According to most accounts, world oil reserves have marched steadily upward over the past 20 years. Extending that apparent trend into the future, one could easily conclude, as the U.S. Energy Information Administration has, that oil production will continue to rise unhindered for decades to come, increasing almost two thirds by 2020.

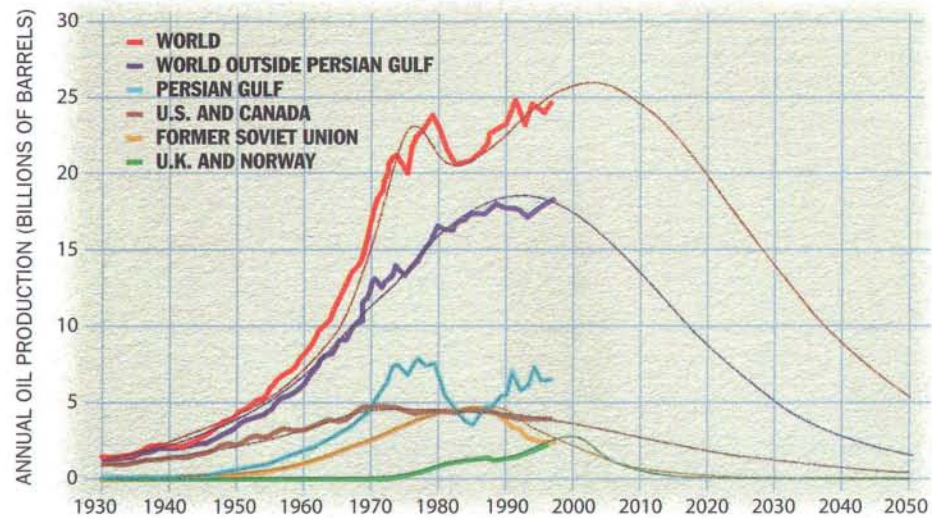
Such growth is an illusion. About 80 percent of the oil produced today flows from fields that were found before 1973, and the great majority of them are declining. In the 1990s oil companies have discovered an average of seven Gbo a year; last year they drained more than three times as much. Yet official figures indicated that proved reserves did not fall by 16 Gbo, as one would expect—rather they *expanded* by 11 Gbo. One reason is that several dozen governments

EARTH'S CONVENTIONAL CRUDE OIL is almost half gone. Reserves (defined here as the amount as likely as not to come out of known fields) and future discoveries together will provide little more than what has already been burned.



LAURIE GRACE

GLOBAL PRODUCTION OF OIL, both conventional and unconventional (*red*), recovered after falling in 1973 and 1979. But a more permanent decline is less than 10 years away, according to the authors' model, based in part on multiple Hubbert curves (*lighter lines*). U.S. and Canadian oil (*brown*) topped out in 1972; production in the former Soviet Union (*yellow*) has fallen 45 percent since 1987. A crest in the oil produced outside the Persian Gulf region (*purple*) now appears imminent.



LAURIE GRACE; SOURCE: JEAN H. LAHERRÈRE

opted not to report declines in their reserves, perhaps to enhance their political cachet and their ability to obtain loans. A more important cause of the expansion lies in revisions: oil companies replaced earlier estimates of the reserves left in many fields with higher numbers. For most purposes, such amendments are harmless, but they seriously distort forecasts extrapolated from published reports.

To judge accurately how much oil explorers will uncover in the future, one has to backdate every revision to the year in which the field was first discovered—not to the year in which a company or country corrected an earlier estimate. Doing so reveals that global discovery peaked in the early 1960s and has been falling steadily ever since. By extending the trend to zero, we can make a good guess at how much oil the industry will ultimately find.

We have used other methods to estimate the ultimate recovery of conventional oil for each country [*see box on next two pages*], and we calculate that the oil industry will be able to recover only about another 1,000 billion barrels of conventional oil. This number, though great, is little more than the 800 billion barrels that have already been extracted.

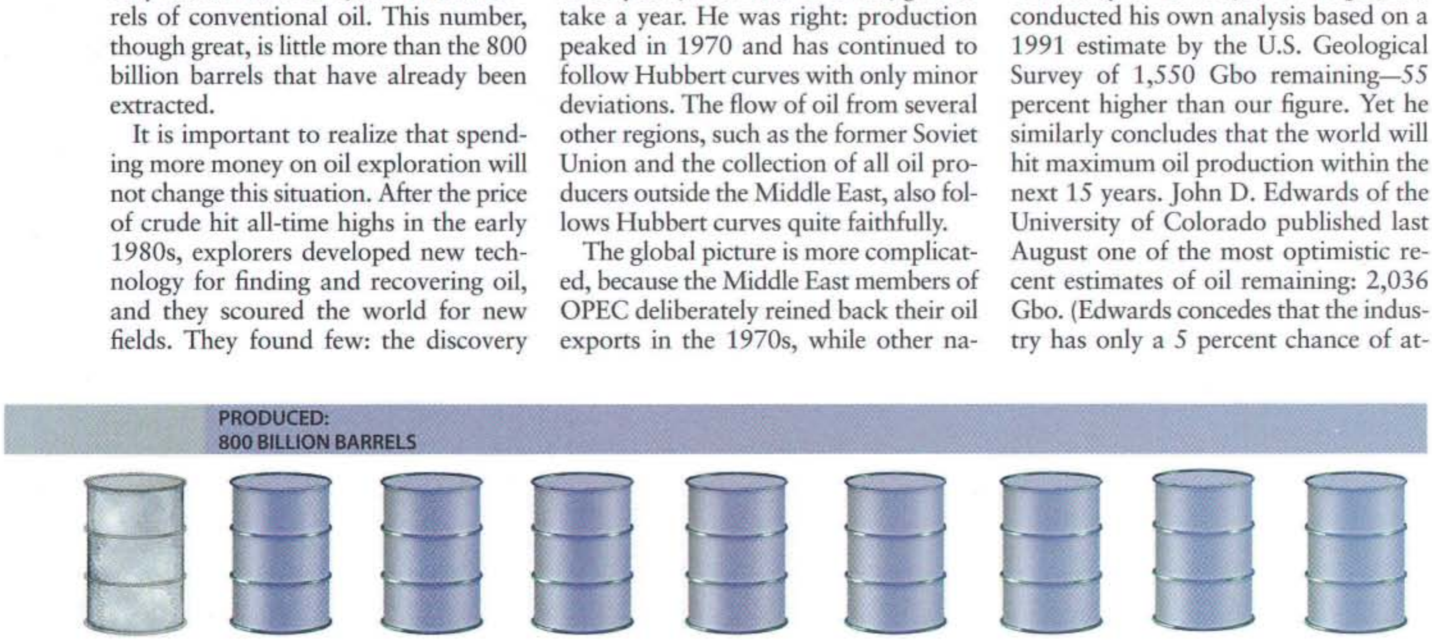
It is important to realize that spending more money on oil exploration will not change this situation. After the price of crude hit all-time highs in the early 1980s, explorers developed new technology for finding and recovering oil, and they scoured the world for new fields. They found few: the discovery rate continued its decline uninterrupted. There is only so much crude oil in the world, and the industry has found about 90 percent of it.

Predicting the Inevitable

Predicting when oil production will stop rising is relatively straightforward once one has a good estimate of how much oil there is left to produce. We simply apply a refinement of a technique first published in 1956 by M. King Hubbert. Hubbert observed that in any large region, unrestrained extraction of a finite resource rises along a bell-shaped curve that peaks when about half the resource is gone. To demonstrate his theory, Hubbert fitted a bell curve to production statistics and projected that crude oil production in the lower 48 U.S. states would rise for 13 more years, then crest in 1969, give or take a year. He was right: production peaked in 1970 and has continued to follow Hubbert curves with only minor deviations. The flow of oil from several other regions, such as the former Soviet Union and the collection of all oil producers outside the Middle East, also follows Hubbert curves quite faithfully.

The global picture is more complicated, because the Middle East members of OPEC deliberately reined back their oil exports in the 1970s, while other nations continued producing at full capacity. Our analysis reveals that a number of the largest producers, including Norway and the U.K., will reach their peaks around the turn of the millennium unless they sharply curtail production. By 2002 or so the world will rely on Middle East nations, particularly five near the Persian Gulf (Iran, Iraq, Kuwait, Saudi Arabia and the United Arab Emirates), to fill in the gap between dwindling supply and growing demand. But once approximately 900 Gbo have been consumed, production must soon begin to fall. Barring a global recession, it seems most likely that world production of conventional oil will peak during the first decade of the 21st century.

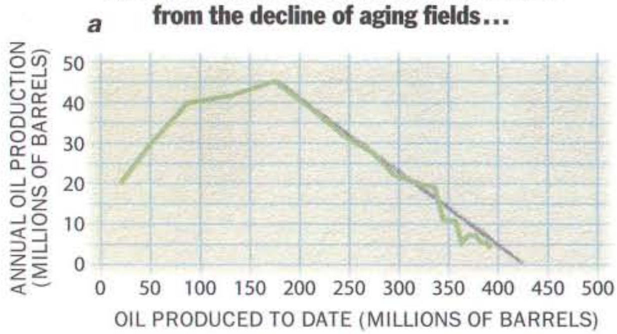
Perhaps surprisingly, that prediction does not shift much even if our estimates are a few hundred billion barrels high or low. Craig Bond Hatfield of the University of Toledo, for example, has conducted his own analysis based on a 1991 estimate by the U.S. Geological Survey of 1,550 Gbo remaining—55 percent higher than our figure. Yet he similarly concludes that the world will hit maximum oil production within the next 15 years. John D. Edwards of the University of Colorado published last August one of the most optimistic recent estimates of oil remaining: 2,036 Gbo. (Edwards concedes that the industry has only a 5 percent chance of at-



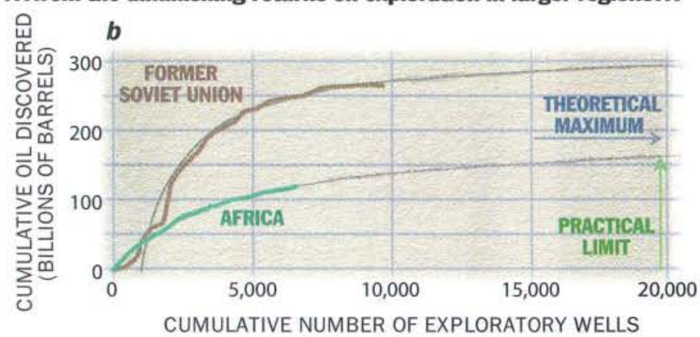
How Much Oil Is Left to Find?

We combined several techniques to conclude that about 1,000 billion barrels of conventional oil remain to be produced. First, we extrapolated published production figures for older oil fields that have begun to decline. The Thistle field off the coast of Britain, for example, will yield about 420 million barrels (*a*). Second, we plotted the amount of oil discovered so far in some regions against the cumulative number of exploratory wells drilled there. Because larger fields tend to be found first—they are simply too large to miss—the curve rises rapidly and then flattens, eventually reaching a theoreti-

We can predict the amount of remaining oil from the decline of aging fields...



...from the diminishing returns on exploration in larger regions...



taining that very high goal.) Even so, his calculations suggest that conventional oil will top out in 2020.

Smoothing the Peak

Factors other than major economic changes could speed or delay the point at which oil production begins to decline. Three in particular have often led economists and academic geologists to dismiss concerns about future oil production with naive optimism.

First, some argue, huge deposits of oil may lie undetected in far-off corners of the globe. In fact, that is very unlikely. Exploration has pushed the frontiers back so far that only extremely deep water and polar regions remain to be fully tested, and even their prospects are now reasonably well understood. Theoretical advances in geochemistry and geophysics have made it possible to map productive and prospective fields with impressive accuracy. As a result, large tracts can be condemned as barren. Much of the deepwater realm, for example, has been shown to be absolutely nonprospective for geologic reasons.

What about the much touted Caspian Sea deposits? Our models project that oil production from that region will grow until around 2010. We agree with analysts at the USGS World Oil Assessment program and elsewhere who rank the total resources there as roughly equivalent to those of the North Sea—that is, perhaps 50 Gbo but certainly not several hundreds of billions as sometimes reported in the media.

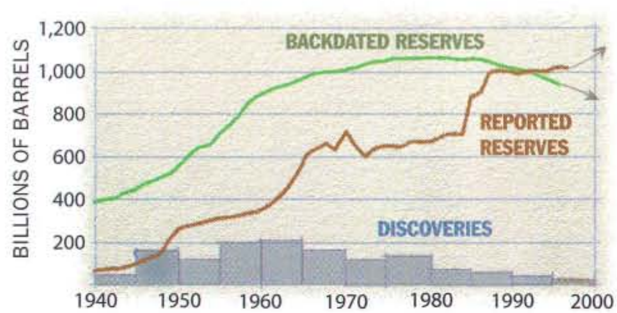
A second common rejoinder is that new technologies have steadily increased the fraction of oil that can be recovered from fields in a basin—the so-called recovery factor. In the 1960s oil companies assumed as a rule of thumb that only 30 percent of the oil in a field was typically recoverable; now they bank on an average of 40 or 50 percent. That progress will continue and will extend global reserves for many years to come, the argument runs.

Of course, advanced technologies will buy a bit more time before production starts to fall [see "Oil Production in the 21st Century," by Roger N. Anderson, on page 68]. But most of the apparent improvement in recovery factors is an artifact of reporting. As oil fields grow old, their owners often deploy newer technology to slow their decline. The falloff also allows engineers to gauge the size of the field more accurately and to correct previous underestimation—in particular P90 estimates that by definition were 90 percent likely to be exceeded.

Another reason not to pin too much hope on better recovery is that oil companies routinely count on technological progress when they compute their reserve estimates. In truth, advanced technologies can offer little help in draining the largest basins of oil, those onshore in the Middle East where the oil needs no assistance to gush from the ground.

Last, economists like to point out that the world contains enormous caches of unconventional oil that can substitute for crude oil as soon as the price rises high enough to make them profitable. There is no question that the resources are ample: the Orinoco oil belt in Venezuela has been assessed to contain a staggering 1.2 trillion barrels of the sludge known as heavy oil. Tar sands and shale deposits in Canada and the former Soviet Union may contain the equivalent of more than 300 billion barrels of oil [see "Mining for Oil," by Richard L. George, on page 66]. Theoretically, these unconventional oil reserves could quench the world's thirst for liquid fuels as conventional oil passes its prime. But the industry will be hard-pressed for the time and money needed to ramp up production of unconventional oil quickly enough.

Such substitutes for crude oil might also exact a high environmental price.

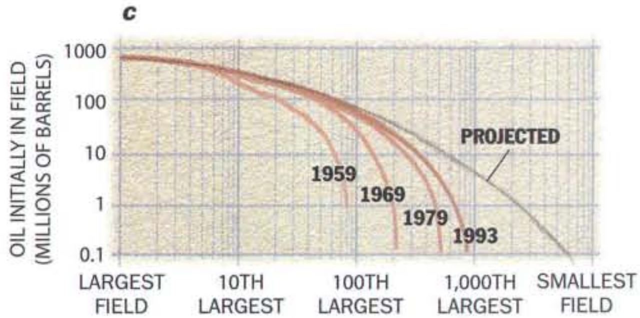


LAURIE GRACE; SOURCE: PETROCONSULTANTS, *OIL AND GAS JOURNAL* AND U.S. GEOLOGICAL SURVEY

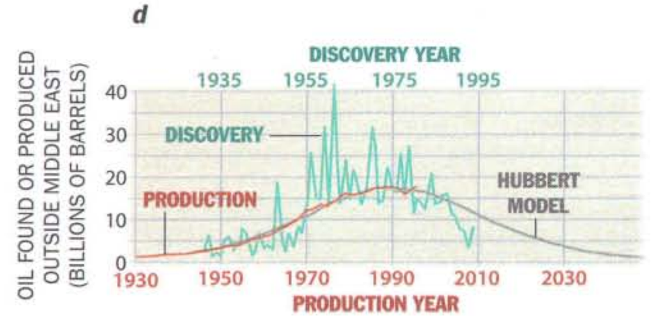
GROWTH IN OIL RESERVES since 1980 is an illusion caused by belated corrections to oil-field estimates. Backdating the revisions to the year in which the fields were discovered reveals that reserves have been falling because of a steady decline in newfound oil (*blue*).

cal maximum: for Africa, 192 Gbo. But the time and cost of exploration impose a more practical limit of perhaps 165 Gbo (*b*). Third, we analyzed the distribution of oil-field sizes in the Gulf of Mexico and other provinces. Ranked according to size and then graphed on a logarithmic scale, the fields tend to fall along a parabola that grows predictably over time (*c*). (Interestingly, galaxies, urban populations and other natural agglomerations also seem to fall along such parabolas.) Finally, we checked our estimates by matching our projections for oil production in large areas, such as the world outside the Persian Gulf region, to the rise and fall of oil discovery in those places decades earlier (*d*). —*C.J.C. and J.H.L.*

...by extrapolating the size of new fields into the future...



...and by matching production to earlier discovery trends.

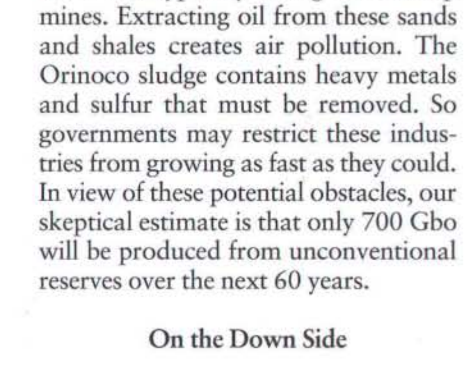


LAURIE GRACE; SOURCE: JEAN H. LAHERRÈRE

Tar sands typically emerge from strip mines. Extracting oil from these sands and shales creates air pollution. The Orinoco sludge contains heavy metals and sulfur that must be removed. So governments may restrict these industries from growing as fast as they could. In view of these potential obstacles, our skeptical estimate is that only 700 Gbo will be produced from unconventional reserves over the next 60 years.

On the Down Side

Meanwhile global demand for oil is currently rising at more than 2 percent a year. Since 1985, energy use is up about 30 percent in Latin America, 40 percent in Africa and 50 percent in Asia. The Energy Information Administration forecasts that worldwide demand for oil will increase 60 percent (to about 40 Gbo a year) by 2020.



SUSPICIOUS JUMP in reserves reported by six OPEC members added 300 billion barrels of oil to official reserve tallies yet followed no major discovery of new fields.

The switch from growth to decline in oil production will thus almost certainly create economic and political tension. Unless alternatives to crude oil quickly prove themselves, the market share of the OPEC states in the Middle East will rise rapidly. Within two years, these nations' share of the global oil business will pass 30 percent, nearing the level reached during the oil-price shocks of the 1970s. By 2010 their share will quite probably hit 50 percent.

The world could thus see radical increases in oil prices. That alone might be sufficient to curb demand, flattening production for perhaps 10 years. (Demand fell more than 10 percent after the 1979 shock and took 17 years to recover.) But by 2010 or so, many Middle Eastern nations will themselves be past the midpoint. World production will then have to fall.

With sufficient preparation, however, the transition to the post-oil economy need not be traumatic. If advanced methods of producing liquid fuels from natural gas can be made profitable and scaled up quickly, gas could become the next source of transportation fuel [see "Liquid Fuels from Natural Gas," by Safaa A. Fouda, on page 74]. Safer nuclear power, cheaper renewable energy, and oil conservation programs could all help postpone the inevitable decline of conventional oil.

Countries should begin planning and investing now. In November a panel of energy experts appointed by President Bill Clinton strongly urged the administration to increase funding for energy research by $1 billion over the next five years. That is a small step in the right direction, one that must be followed by giant leaps from the private sector.

The world is not running out of oil—at least not yet. What our society does face, and soon, is the end of the abundant and cheap oil on which all industrial nations depend. SA

The Authors

COLIN J. CAMPBELL and JEAN H. LAHERRÈRE have each worked in the oil industry for more than 40 years. After completing his Ph.D. in geology at the University of Oxford, Campbell worked for Texaco as an exploration geologist and then at Amoco as chief geologist for Ecuador. His decade-long study of global oil-production trends has led to two books and numerous papers. Laherrère's early work on seismic refraction surveys contributed to the discovery of Africa's largest oil field. At Total, a French oil company, he supervised exploration techniques worldwide. Both Campbell and Laherrère are currently associated with Petroconsultants in Geneva.

Further Reading

UPDATED HUBBERT CURVES ANALYZE WORLD OIL SUPPLY. L. F. Ivanhoe in *World Oil,* Vol. 217, No. 11, pages 91–94; November 1996.

THE COMING OIL CRISIS. Colin J. Campbell. Multi-Science Publishing and Petroconsultants, Brentwood, England, 1997.

OIL BACK ON THE GLOBAL AGENDA. Craig Bond Hatfield in *Nature,* Vol. 387, page 121; May 8, 1997.

Mining for Oil


COURTESY OF SYNCRUDE

More oil is trapped in Canadian sands than Saudi Arabia holds in its reserves. The technology now exists to exploit this vast resource profitably

by Richard L. George

The term "oil" has, to date, been synonymous with conventional crude oil, a liquid mixture of hydrocarbons that percolates through porous strata and flows readily up drilled boreholes. But much of the world's remaining endowment of oil takes a less convenient form: a black, tarlike substance called bitumen, which sticks stubbornly in the pore spaces between the grains of certain sands and shales (solidified muds). Because bitumen normally will not flow through such formations, the straightforward way to recover it is to scoop it out of open-pit mines.

Digging for oil is certainly more troublesome than simply drilling and pumping, but the enormity of this resource makes it hard to ignore. Current processing methods could recover about 300 billion barrels from oil sands in the Canadian province of Alberta alone—more than the reserves of conventional oil in Saudi Arabia. Oil shale appears to offer a less prodigious supply, but Australia contains at least 28 billion barrels of petroleum in this form, and other deposits lie buried in Estonia, Brazil, Sweden, the U.S. and China. All told, oil sands and shales around the world could, in principle, hold several trillion barrels of oil.

Yet it is difficult to predict how much of that potential can be profitably recovered, because the processing needed to turn oil sands or shales into useful petroleum products is quite challenging. My company, Suncor Energy, is one of only two in the world that have successfully exploited oil sands by mining them.

The roots of our enterprise in northern Alberta go back many years. For centuries, natives living in the region

The Processing of Oil Sands

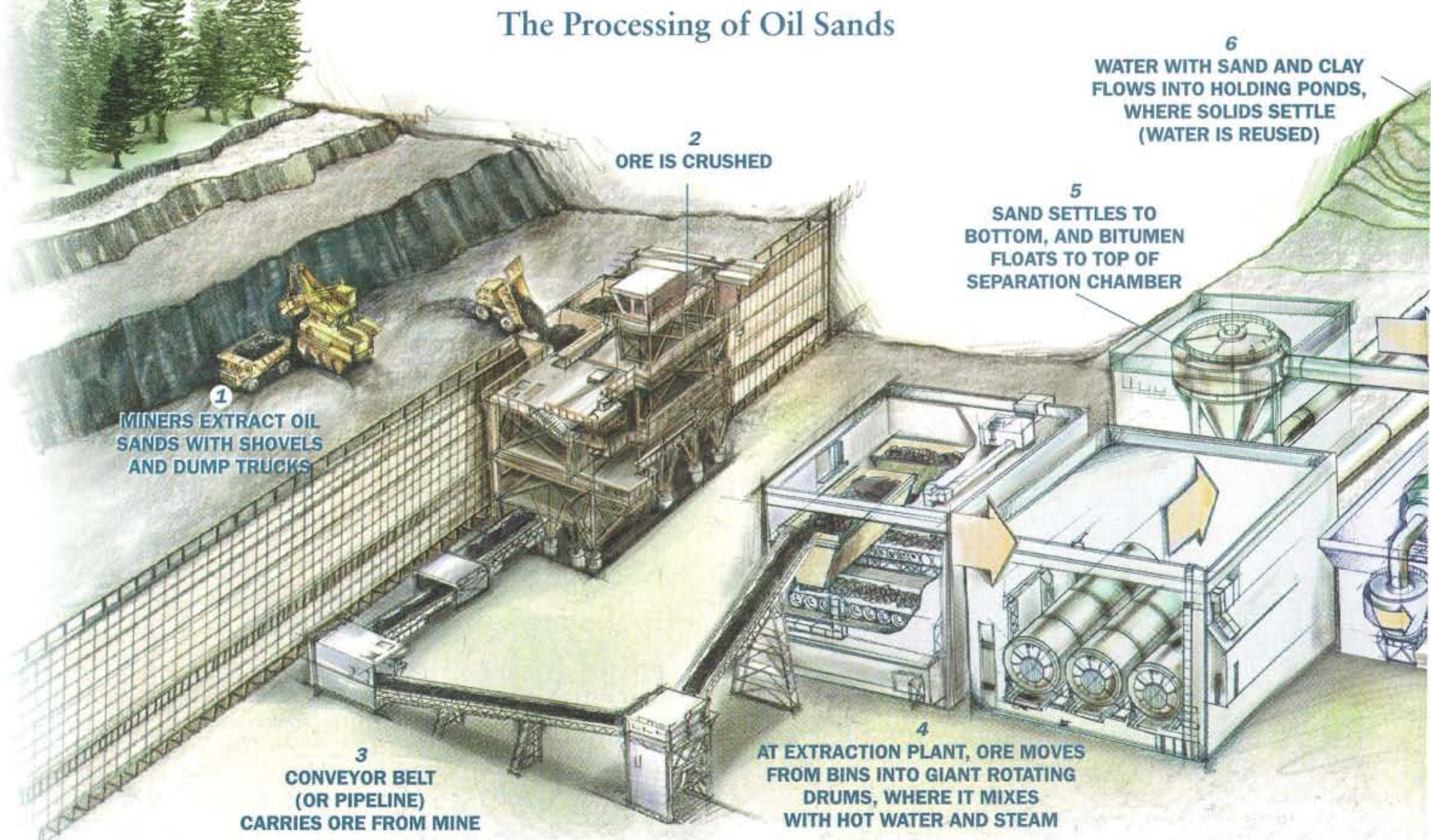


ILLUSTRATION BY DANIELS & DANIELS; SOURCE: SUNCOR

OIL SANDS, coated with tarlike bitumen, have the appearance of coffee grounds.

used the sticky bitumen that oozes out of the banks of the Athabasca River to patch leaks in their canoes. And as early as 1893 the Canadian government sponsored investigations of the Athabasca "tar sands" as a potential source of petroleum. Then, in 1920, Karl A. Clark of the Alberta Research Council found a practical way to separate the bitumen from the sand. After shoveling some of the asphalt into the family washing machine and adding hot water and caustic soda, he discovered that the bitumen floated to the surface as a frothy foam, ready to be skimmed off.

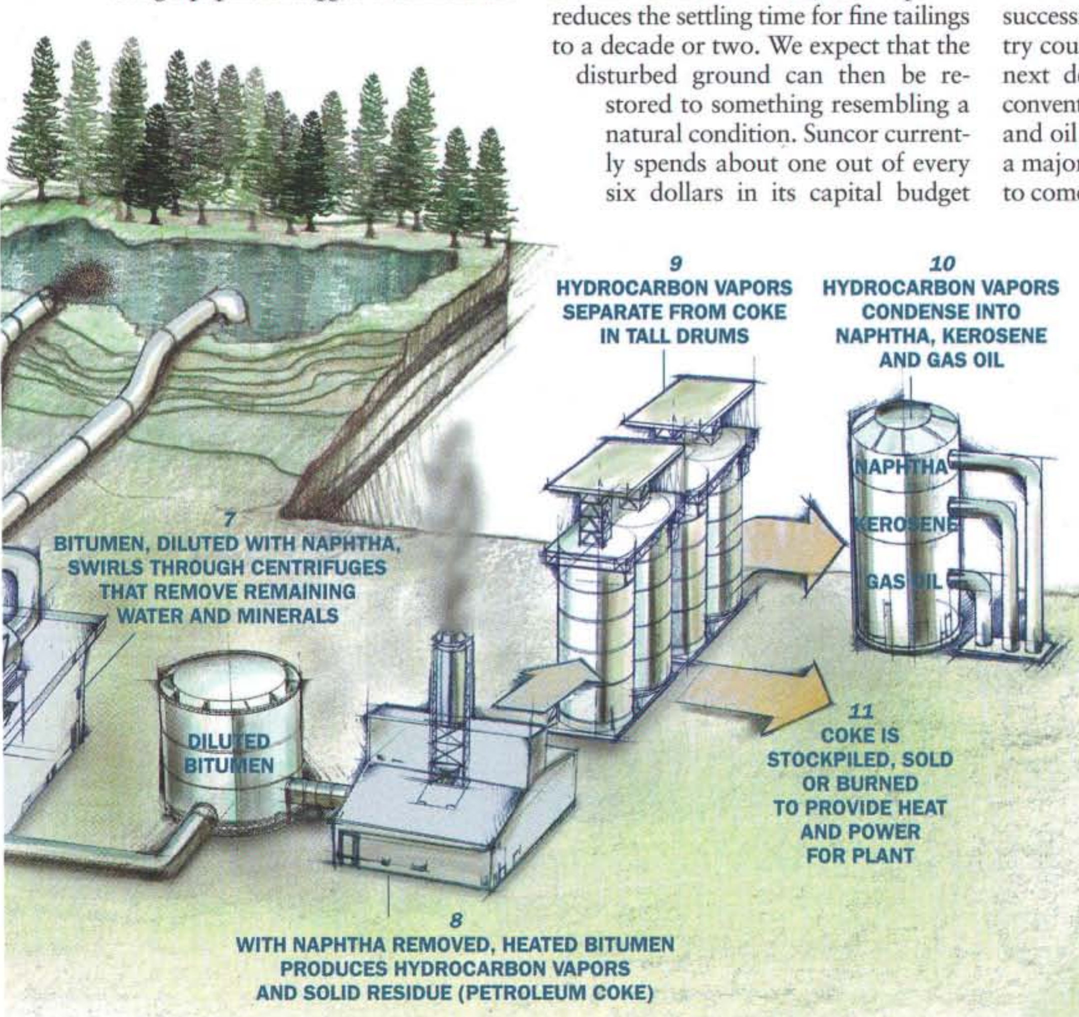
Clark's method was clearly workable. Yet the idea languished for decades until the precursor of Suncor Energy, the Great Canadian Oil Sands Ltd., began large-scale mining of oil sands in 1967. Rising petroleum prices during the 1970s helped to keep the expensive operation afloat. But failures of the excavating equipment dogged the miners until 1992, when Suncor modernized the facility in a concerted effort to reduce the cost of extracting this oil.

For the past five years, my colleagues at Suncor have been producing oil that is quite profitable at current market prices. Our production has expanded by 38 percent over that period, and we now sell 28 million barrels of unconventional oil a year. This growth will most likely accelerate in the future, limited in part by the need to ensure that the local environment is harmed as little as possible.

The handling of residues, called tailings, is of particular concern. The coarser grains settle quickly from a slurry of water and sand, and we put these materials back into the ground. But the water still contains many fine particles, so we have to hold it in large ponds to avoid contaminating nearby rivers and streams. Left alone, the fine tailings would not sink to the bottom for centuries. But researchers in industry and government discovered that adding gypsum (a by-product of the sulfur that is removed from the oil) to the ponds reduces the settling time for fine tailings to a decade or two. We expect that the disturbed ground can then be restored to something resembling a natural condition. Suncor currently spends about one out of every six dollars in its capital budget on reducing environmental disturbance.

One alternative technology for extracting bitumen sidesteps such problems and could, in principle, allow industry to tap huge deposits that are too deep to mine. It turns out that the oil in these sands can be made to flow by injecting steam into the ground. Once it is heated, the oil thins and pools underneath the site of injection. Ordinary oilfield equipment can then bring it to the surface. This process, called steam-assisted gravity drainage, is now being tested by oil companies such as the Alberta Energy Company. Suncor, too, may exploit this approach in the future.

Engineers at Suncor are also examining a system to bake the oil out of crushed rock in a giant, drum-shaped kiln. Although this method (invented by William Taciuk, collaborating with the Alberta Department of Energy) is not particularly appropriate for oil sands, it appears to work well for processing oil shales. If a demonstration plant that Suncor is building with its Australian partners—Southern Pacific Petroleum and Central Pacific Minerals—proves successful, a whole new oil shale industry could develop in Australia over the next decade. So as production from conventional oil fields dwindles, oil shale and oil sand reserves may well become a major source of energy in the century to come.



The Author

RICHARD L. GEORGE is president and chief executive officer of Suncor Energy, which now mines oil sands in Fort McMurray, Alberta, and plans to develop oil shales in Queensland, Australia. George has a bachelor of science degree in engineering from Colorado State University and a law degree from the University of Houston.

Further Reading

CANADA'S OIL SANDS: THE UNCONVENTIONAL ALTERNATIVES. Richard L. George in *Hart's Petroleum Engineer International,* pages 33–36; May 1997.

Suncor Energy World Wide Web site: http://www.suncor.com

Syncrude World Wide Web site: http://www.syncrude.com

Oil Production in the 21st Century

Recent innovations in underground imaging, steerable drilling and deepwater oil production could recover more of what lies below

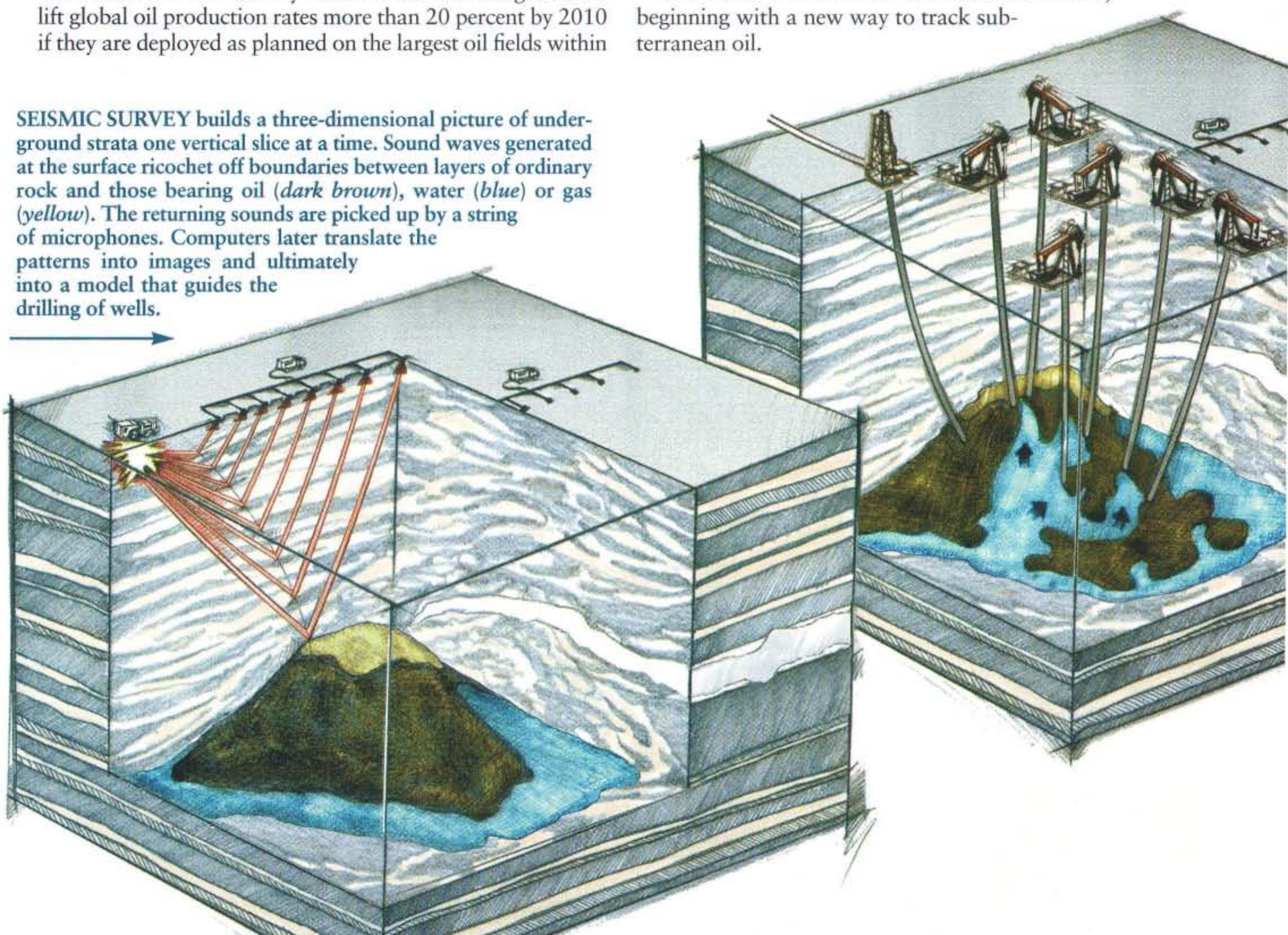
by Roger N. Anderson

On the face of it, the outlook for conventional oil—the cheap, easily recovered crude that has furnished more than 95 percent of all oil to date—seems grim. In 2010, according to forecasts, the world's oil-thirsty economies will demand about 10 billion more barrels than the industry will be able to produce. A supply shortfall that large, equal to almost half of all the oil extracted in 1997, could lead to price shocks, economic recession and even wars.

Fortunately, four major technological advances are ready to fill much of the gap by accelerating the discovery of new oil reservoirs and by dramatically increasing the fraction of oil within existing fields that can be removed economically, a ratio known as the recovery factor. These technologies could lift global oil production rates more than 20 percent by 2010 if they are deployed as planned on the largest oil fields within three to five years. Such rapid adoption may seem ambitious for an industry that traditionally has taken 10 to 20 years to put new inventions to use. But in this case, change will be spurred by formidable economic forces.

For example, in the past two years, the French oil company Elf has discovered giant deposits off the coast of West Africa. In the same period, the company's stock doubled, as industry analysts forecasted that Elf's production would increase by 8 percent in 2001. If the other major oil producers follow suit, they should be able by 2010 to provide an extra five billion barrels of oil each year, closing perhaps half the gap between global supply and demand.

This article will cover the four advances in turn, beginning with a new way to track subterranean oil.



SEISMIC SURVEY builds a three-dimensional picture of underground strata one vertical slice at a time. Sound waves generated at the surface ricochet off boundaries between layers of ordinary rock and those bearing oil (*dark brown*), water (*blue*) or gas (*yellow*). The returning sounds are picked up by a string of microphones. Computers later translate the patterns into images and ultimately into a model that guides the drilling of wells.

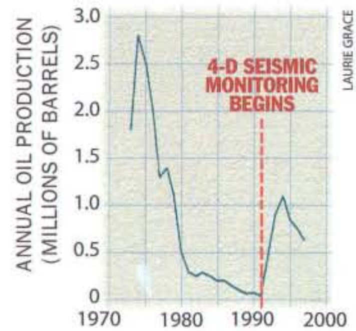
DANIELS & DANIELS

Tracking Oil in Four Dimensions

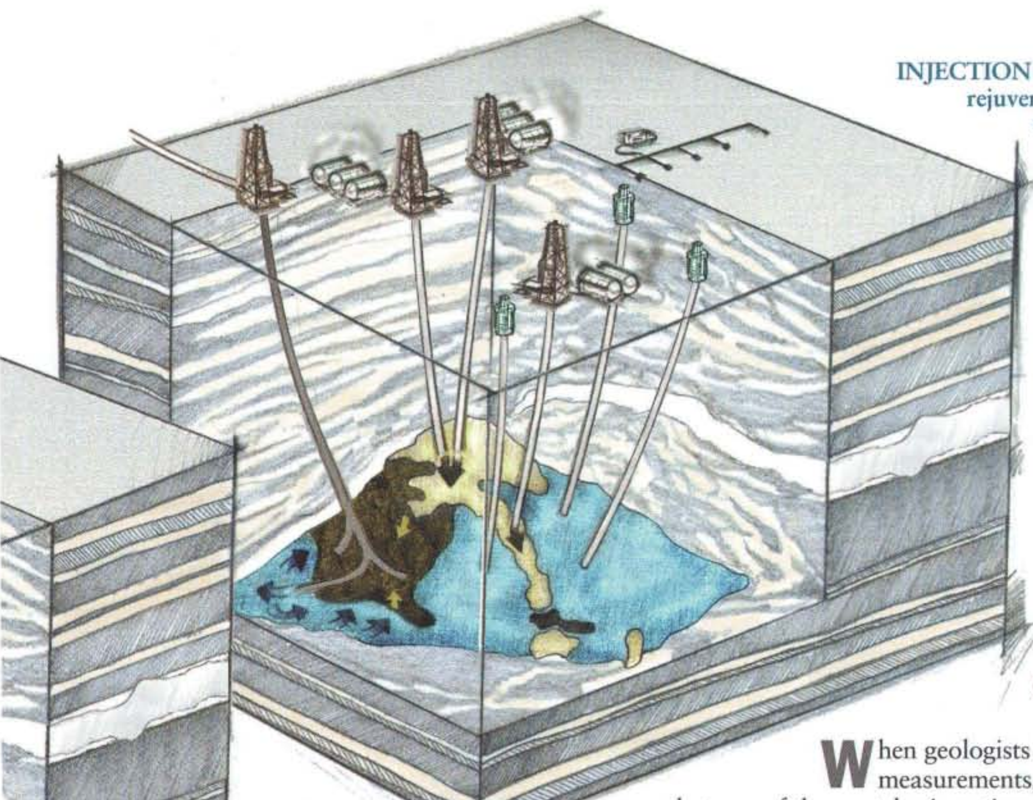
Finding oil became much more efficient after 1927, when geologists first successfully translated acoustic reflections into detailed cross sections of the earth's crust. Seismologists later learned how to piece together several such snapshots to create three-dimensional models of the oil locked inside layers of porous rock. Although this technique, known as 3-D seismic analysis, took more than a decade to become standard practice, it is now credited with increasing oil discovery and recovery rates by 20 percent.

In recent years, scientists in my laboratory at Columbia University and elsewhere have developed even more powerful techniques capable of tracking the movement of oil, gas and water as drilled wells drain the subterranean strata—a "4-D" scheme that includes the added dimension of time. This information can then be used to do a "what if" analysis on the oil field, designing ways to extract as much of the oil as quickly and cheaply as possible.

Compared with its predecessor, the 4-D approach seems to be catching on quickly: the number of oil fields benefiting from it has doubled in each of the past four years and now stands at about 60. Such monitoring can boost recovery factors by 10 to 15 percentage points. Unfortunately, the technique will work in only about half the world's major fields, those where relatively soft rock is suffused with oil and natural gas.



FLOW OF OIL from a reservoir in the largest field off the Louisiana shore resurged in 1992, shortly after operators began using 4-D seismic monitoring to locate hidden caches of oil.



INJECTION OF LIQUID CARBON DIOXIDE can rejuvenate dying oil fields. Pumped at high pressure from tanks into wells that have ceased producing oil, the carbon dioxide flows through the reservoir and, if all goes well, pushes the remaining oil down toward active wells. Steam and natural gas are sometimes also used for this purpose. Alternatively, water can be injected below a pocket of bypassed crude in order to shepherd the oil into a well. In the future, "smart" wells currently under development will be able to retrieve oil simultaneously from some branches of the well while using other branches to pump water out of the oil stream and back into the formation from which it came.

PRODUCTION WELLS often draw water from below and gas from above into pore spaces once full of oil. This complex flow strands pockets of crude far from wells; traditional drilling techniques thus miss up to two thirds of the oil in a reservoir. But repeated seismic surveys can now be assembled into a 4-D model that not only tracks where oil, gas and water in the field are located but also predicts where they will go next. Advanced seismic monitoring works well on about half the world's oil fields, but it fails on oil buried in very hard rock or beneath beds of salt (*thick white layer*).

Gassing Things Up

When geologists began studying the new time-lapse measurements, they were surprised to discover that one of the most basic notions about oil movement—that it naturally settles between lighter gas above and heavier groundwater below—oversimplifies the behavior of real oil fields. In fact, most wells produce complex, fractal drainage patterns that cause the oil to mix with gas and water. As a result, specialists now know that the traditional technique of pumping a well until the oil slows to a trickle often leaves 60 percent or more of the oil behind.

A more efficient strategy is to pump natural gas, steam or liquid carbon dioxide into dead wells. The infusion then spreads downward through pores in the rock and, if one has planned carefully, pushes oil that otherwise would have been abandoned toward a neighboring well. Alternatively, water is often pumped below the oil to increase its pressure, helping it flow up to the surface.

Injections of steam and carbon dioxide have been shown to increase recovery factors by 10 to 15 percentage points. Unfortunately, they also raise the cost of oil production by 50 to 100 percent—and that added expense falls on top of a 10 to 25 percent surcharge for 4-D seismic monitoring. So unless carbon dioxide becomes much cheaper (perhaps because global-warming treaties restrict its release) these techniques will probably continue to serve only as a last resort.

Steering to Missed Oil

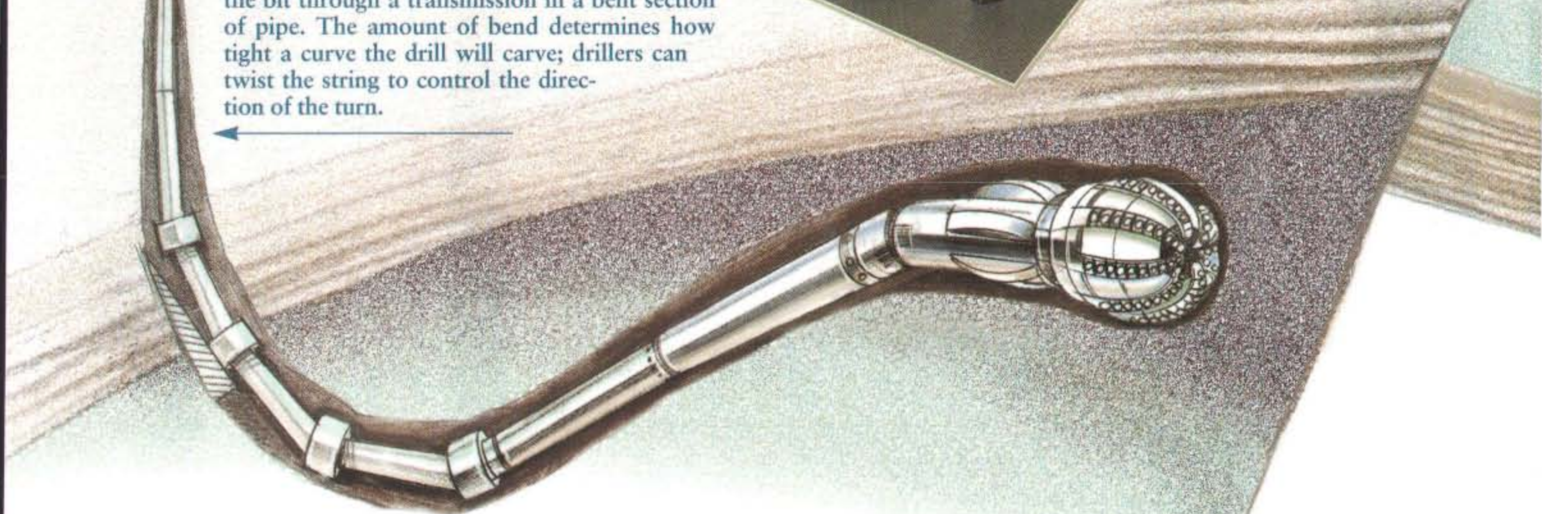
A third major technological advance, known as directional drilling, can tap bypassed deposits of oil at less expense than injection. Petroleum engineers can use a variety of new equipment to swing a well from vertical to entirely horizontal within a reservoir several kilometers underground.

Traditionally, drillers rotated the long steel pipe, or "string," that connects the rig at the surface to the bit at the bottom of the well. That method fails when the pipe must turn a corner—the bend would break the rotating string. So steerable drill strings do not rotate; instead a mud-driven motor inserted near the bit turns only the diamond-tipped teeth that do the digging. An elbow of pipe placed between the mud motor and the bit controls the direction of drilling.

Threading a hole through kilometers of rock into a typical oil zone 30 meters (about 100 feet) thick is precise work. Schlumberger, Halliburton and other international companies have developed sophisticated sensors that significantly improve the accuracy of drilling. These devices, which operate at depths of up to 6,000 meters and at temperatures as high as 200 degrees Celsius (400 degrees Fahrenheit), attach to the drill pipe just above or below the mud motor. Some measure the electrical resistance of the surrounding rock. Others send out neutrons and gamma rays; then they count the number that are scattered back by the rock and pore fluids. These measurements and the current position of the bit (calculated by an inertial guidance system) are sent back to the surface through pulses in the flow of the very mud used to turn the motor and lubricate the well bore. Engineers can adjust the path of the drill accordingly, thus snaking their way to the most oil-rich part of the formation.

Once the hole is completed, drillers typically erect production equipment on top of the wellhead. But several companies are now developing sensors that can detect the mix of oil, gas and water near its point of entry deep within the well. "Smart" wells with such equipment will be able to separate water out of the well stream so that it never goes to the surface. Instead a pump, controlled by a computer in the drill pipe, will inject the wastewater below the oil level.

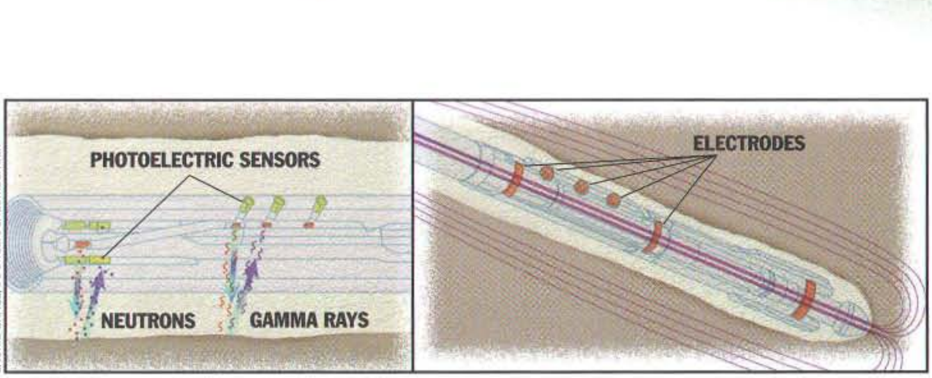
DANIELS & DANIELS (background illustration)



HORIZONTAL DRILLING was impractical when oil rigs had to rotate the entire drill string—up to 5,800 meters (roughly 19,000 feet) of it—in order to turn the rock-cutting bit at the bottom. Wells that swing 90 degrees over a space of just 100 meters are now common thanks to the development of motors that can run deep underground. The motor's driveshaft connects to the bit through a transmission in a bent section of pipe. The amount of bend determines how tight a curve the drill will carve; drillers can twist the string to control the direction of the turn.

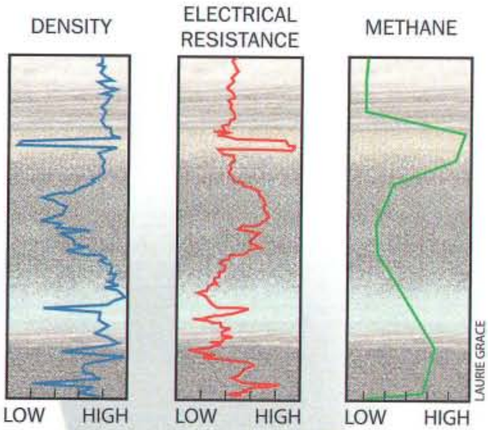
HITEC DRILLING & MARINE SYSTEMS

DRILLING CONSOLE allows an engineer at the surface to monitor sensors near the drill bit that indicate whether it has hit oil or water. The drill can then be steered into position for the optimum yield.

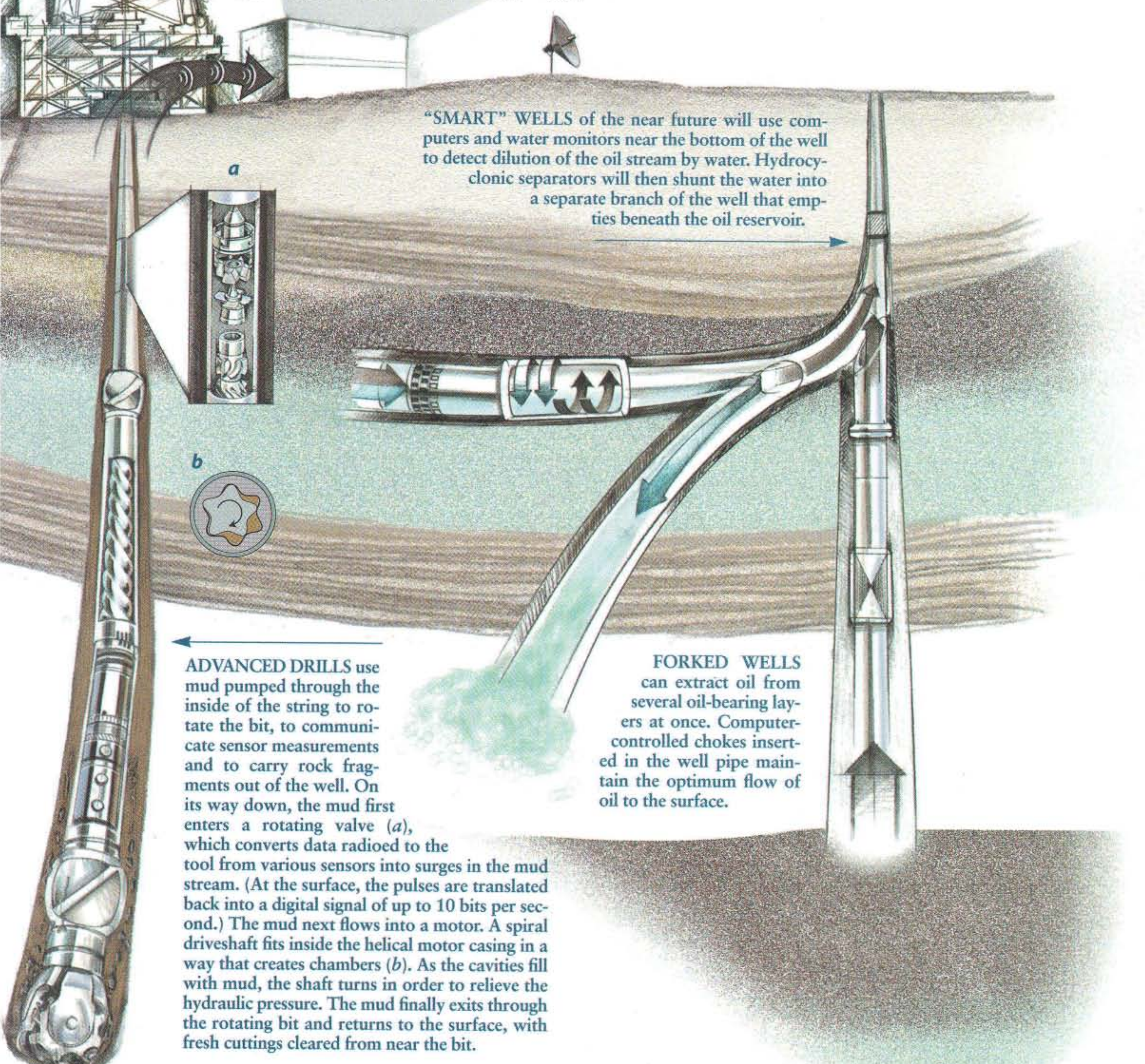


LAURIE GRACE; SOURCE: SCHLUMBERGER

SENSORS near the bit can detect oil, water and gas. One device measures the porosity of the surrounding rock by emitting neutrons, which scatter off hydrogen atoms. Another takes a density reading by shooting out gamma rays that interact with adjacent electrons. Oil and water also affect electrical resistance, measured from a current passed through the bit, the rock and nearby electrodes.



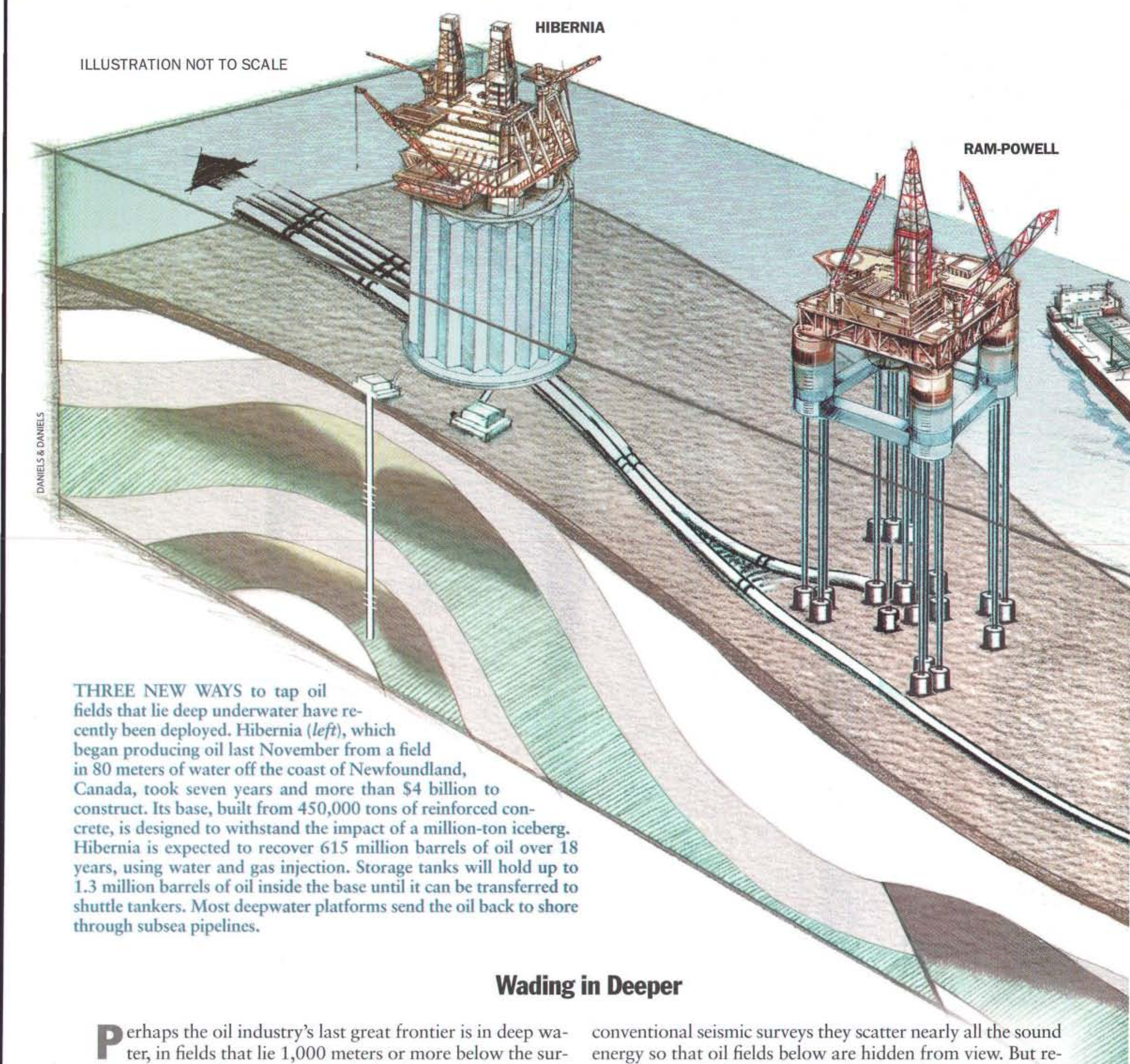
GEOLOGIC MEASUREMENTS collected by sensors near the bottom of the drill pipe can be analyzed at the wellhead or transmitted via satellite to engineers anywhere in the world. Several characteristics of the rocks surrounding the drill bit can reveal the presence of oil or gas (*left*). Petroleum tends to accumulate in relatively light, porous rocks, for example, so some geosteering systems calculate the bulk density of nearby strata. Others measure the electrical resistance of the earth around the drill; layers soaked with briny water have a much lower resistance than those rich in oil. Gas chromatographs at the surface analyze the returning flow of lubricating mud for natural gas captured during its journey.



"SMART" WELLS of the near future will use computers and water monitors near the bottom of the well to detect dilution of the oil stream by water. Hydrocyclonic separators will then shunt the water into a separate branch of the well that empties beneath the oil reservoir.

ADVANCED DRILLS use mud pumped through the inside of the string to rotate the bit, to communicate sensor measurements and to carry rock fragments out of the well. On its way down, the mud first enters a rotating valve (*a*), which converts data radioed to the tool from various sensors into surges in the mud stream. (At the surface, the pulses are translated back into a digital signal of up to 10 bits per second.) The mud next flows into a motor. A spiral driveshaft fits inside the helical motor casing in a way that creates chambers (*b*). As the cavities fill with mud, the shaft turns in order to relieve the hydraulic pressure. The mud finally exits through the rotating bit and returns to the surface, with fresh cuttings cleared from near the bit.

FORKED WELLS can extract oil from several oil-bearing layers at once. Computer-controlled chokes inserted in the well pipe maintain the optimum flow of oil to the surface.



THREE NEW WAYS to tap oil fields that lie deep underwater have recently been deployed. Hibernia (*left*), which began producing oil last November from a field in 80 meters of water off the coast of Newfoundland, Canada, took seven years and more than $4 billion to construct. Its base, built from 450,000 tons of reinforced concrete, is designed to withstand the impact of a million-ton iceberg. Hibernia is expected to recover 615 million barrels of oil over 18 years, using water and gas injection. Storage tanks will hold up to 1.3 million barrels of oil inside the base until it can be transferred to shuttle tankers. Most deepwater platforms send the oil back to shore through subsea pipelines.

Wading in Deeper

Perhaps the oil industry's last great frontier is in deep water, in fields that lie 1,000 meters or more below the surface of the sea. Petroleum at such depths used to be beyond reach, but no longer. Remotely controlled robot submarines can now install on the seafloor the complex equipment needed to guard against blowouts, to regulate the flow of oil at the prevailing high pressures and to prevent natural gas from freezing and plugging pipelines. Subsea complexes will link clusters of horizontal wells. The collected oil will then be funneled both to tankers directly above and to existing platforms in shallower waters through long underwater pipelines. In just the next three years, such seafloor facilities are scheduled for construction in the Gulf of Mexico and off the shores of Norway, Brazil and West Africa.

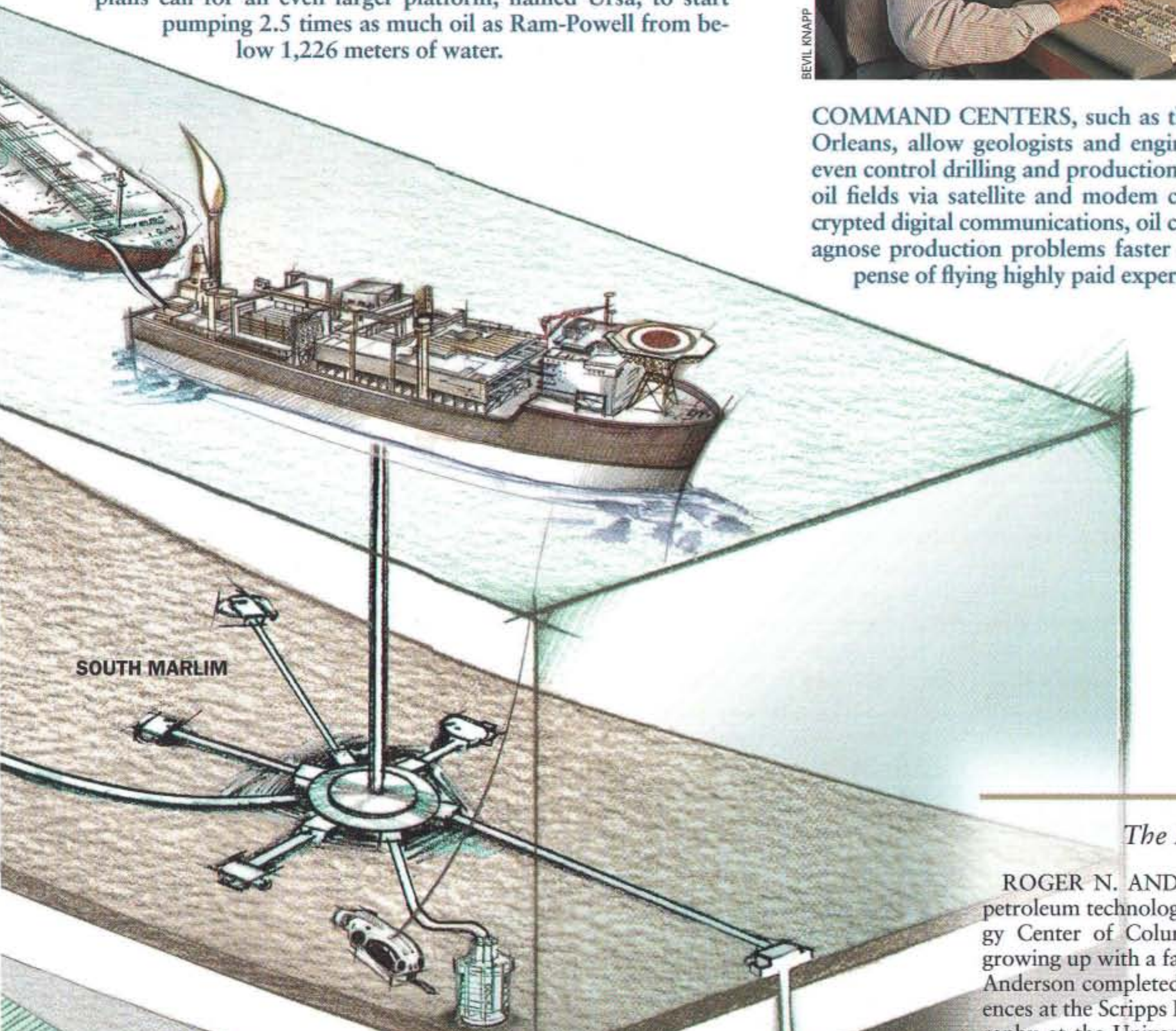
More than deep water alone hinders the exploitation of offshore oil and gas fields. Large horizontal sheets of salt and basalt (an igneous rock) sometimes lie just underneath the seafloor in the deep waters of the continental margins. In conventional seismic surveys they scatter nearly all the sound energy so that oil fields below are hidden from view. But recently declassified U.S. Navy technology for measuring tiny variations in the force and direction of gravity, combined with ever expanding supercomputer capabilities, now allows geophysicists to see under these blankets of salt or basalt.

Extracting oil from beneath the deep ocean is still enormously expensive, but innovation and necessity have led to a new wave of exploration in that realm. Already the 10 largest oil companies working in deep water have discovered new fields that will add 5 percent to their combined oil reserves, an increase not yet reflected in global reserve estimates.

The technology for oil exploration and production will continue to march forward in the 21st century. Although it is unlikely that these techniques will entirely eliminate the impending shortfall in the supply of crude oil, they will buy critical time for making an orderly transition to a world fueled by other energy sources.

RAM-POWELL platform (*center*), built by Shell Oil, Amoco and Exxon, began production in the Gulf of Mexico last September. The 46-story platform is anchored to 270-ton piles driven into the seafloor 980 meters below. Twelve tendons, each 71 centimeters in diameter, provide a strong enough mooring to withstand 22-meter waves and hurricane winds up to 225 kilometers per hour. The $1-billion facility can sink wells up to six kilometers into the seabed in order to tap the 125 million barrels of recoverable oil estimated to lie in the field. A 30-centimeter pipeline will transport the oil to platforms in shallower water 40 kilometers away. Ram-Powell is the third such tension leg platform completed by Shell in three years. Next year, Shell's plans call for an even larger platform, named Ursa, to start pumping 2.5 times as much oil as Ram-Powell from below 1,226 meters of water.

COMMAND CENTERS, such as the one above in New Orleans, allow geologists and engineers to monitor and even control drilling and production equipment in remote oil fields via satellite and modem connections. With encrypted digital communications, oil companies can now diagnose production problems faster and can save the expense of flying highly paid experts around the world.



DEEPEST OIL WELL in active production (*above*) currently lies more than 1,709 meters beneath the waves of the South Atlantic Ocean, in the Marlim field off the coast of Campos, Brazil. The southern part of this field alone is thought to contain 10.6 billion barrels of oil. Such resources were out of reach until recently. Now remotely operated submarines are being used to construct production facilities on the sea bottom itself. The oil can then be piped to a shallower platform if one is nearby. Or, as in the case of the record-holding South Marlim 3B well, a ship can store the oil until shuttle tankers arrive. The challenge is to hold the ship steady above the well. Moorings can provide stability at depths up to about 1,500 meters. Beyond that limit, ships may have to use automatic thrusters linked to the Global Positioning System and beacons on the seafloor to actively maintain their position. These techniques may allow the industry to exploit oil fields under more than 3,000 meters of water in the near future.

The Author

ROGER N. ANDERSON is director of petroleum technology research at the Energy Center of Columbia University. After growing up with a father in the oil industry, Anderson completed his Ph.D. in earth sciences at the Scripps Institution of Oceanography at the University of California, San Diego. He sits on the board of directors of Bell Geospace and 4-D Systems and spends his summers consulting for oil and service companies. Anderson has published more than 150 peer-reviewed scientific papers and holds seven U.S. patents.

Further Reading

DIRECTIONAL DRILLING. George A. Cooper in *Scientific American,* Vol. 270, No. 5, May 1994.

DEEPWATER TECHNOLOGY. Gulf Publishing. Supplement to *World Oil,* August 1997.

WORLD OIL'S 4-D SEISMIC HANDBOOK: THE FOURTH DIMENSION IN RESERVOIR MANAGEMENT. Gulf Publishing, 1997.

Liquid Fuels from Natural Gas

Natural gas is cleaner and more plentiful than oil. New ways to convert it to liquid form may soon make it just as cheap and convenient to use in vehicles

by Safaa A. Fouda

Recently countless California motorists have begun contributing to a remarkable transition. Few of these drivers realize that they are doing something special when they tank up their diesel vehicles at the filling station. But, in fact, they are helping to wean America from crude oil by buying a fuel made in part from natural gas.

Diesel fuel produced in this unconventional way is on sale in California because the gas from which it is derived is largely free of sulfur, nitrogen and heavy metals—substances that leave the tailpipe as noxious pollutants. Blends of ordinary diesel fuel and diesel synthesized from natural gas (currently produced commercially by Shell in Indonesia) meet the toughest emissions standards imposed by the California Air Resources Board.

But natural gas is not only the cleanest of fossil fuels, it is also one of the most plentiful. Industry analysts estimate that the world holds enough readily recoverable natural gas to produce 500 billion barrels of synthetic crude—more than twice the amount of oil ever found in the U.S. Perhaps double that quantity of gas can be found in coal seams and in formations that release gas only slowly. Thus, liquid fuels derived from natural gas could keep overall production on the rise for about a decade after conventional supplies of crude oil begin to dwindle.

Although global stocks of natural gas are enormous, many of the deposits lie far from the people in need of energy. Yet sending gas over long distances often turns out to be prohibitively expensive. Natural gas costs four times as much as crude oil to transport through pipelines because it has a much lower energy density. The so-called stranded gas can be cooled and compressed into a liquid for shipping by tanker. Unfortunately, the conversion facilities required are large and complex, and because liquefied natural gas is hard to handle, the demand for it is rather limited.

But what if there were a cheap way to convert natural gas to a form that remains liquid at room temperature and pressure? Doing so would allow the energy to be piped to markets inexpensively. If the liquid happened to be a fuel that worked in existing vehicles, it could substitute for oil-based gasoline and diesel. And oil producers would stand to profit in many instances by selling liquid fuels or other valuable chemicals made using the gas coming from their wells.

Right now the gas released from oil wells in many parts of the world holds so little value that it is either burned on site or reinjected into the ground. In Alaska alone, oil companies pump about 200 million cubic meters (roughly seven billion cubic feet) of natural gas back into the ground daily—in large part to avoid burdening the atmosphere with additional carbon dioxide, a worrisome greenhouse gas.

But recent technical advances have prompted several oil companies to consider building plants to convert this natural gas into liquid form, which could then be delivered economically through the Alaska pipeline. On the Arabian Peninsula, the nation of Qatar is negotiating with three petrochemical companies to build gas conversion plants that would exploit a huge offshore field—a single reservoir that contains about a tenth of the world's proved gas reserves. And Norway's largest oil company, Statoil, is looking at building relatively small modules mounted on floating platforms to transform gas in remote North Sea fields into liquids. Although these efforts will use somewhat different technologies, they all must address the same fundamental problem in chemistry: making larger hydrocarbon molecules from smaller ones.

The Classic Formula

The main component of natural gas is methane, a simple molecule that has four hydrogen atoms neatly arrayed around one carbon atom. This symmetry makes methane particularly stable. Converting it to a liquid fuel requires first breaking its chemical bonds. High temperatures and pressures help to tear these bonds apart. So do cleverly designed catalysts, substances that can foster a chemical reaction without themselves being consumed.

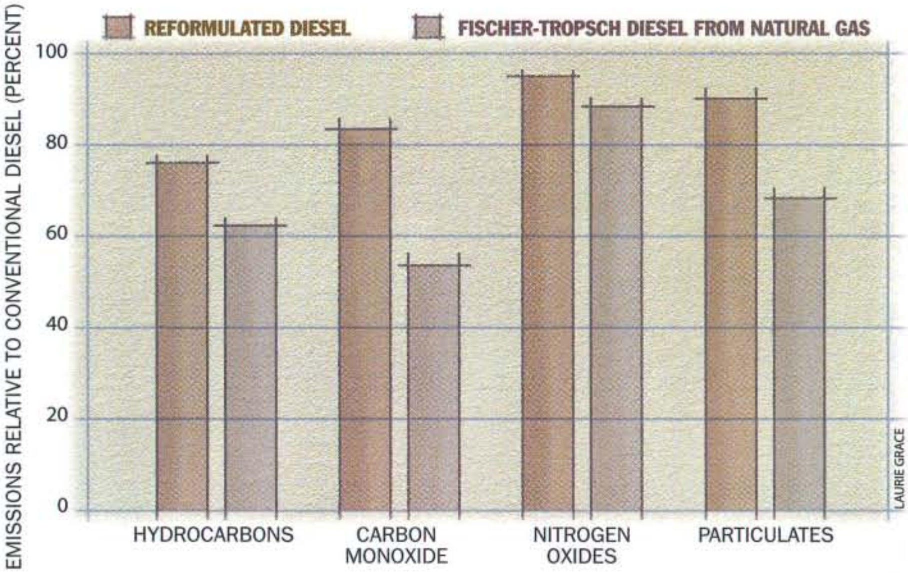
The conventional "indirect" approach for converting natural gas to liquid form relies on brute force. First, the chemical bonds in methane are broken using steam, heat and a nickel-based catalyst to produce a mixture of carbon monoxide and hydrogen known as syngas (or, more formally, synthesis gas). This process is called steam re-forming.

The second step in the production of liquid fuels (or other valuable petrochemicals) from syngas uses a method invented in 1923 by Franz Fischer and Hans Tropsch. During World War II, Germany harnessed this technique to produce liquid fuels using syngas made from coal and atmospheric oxygen, thus establishing a reliable internal source for gasoline and diesel.

This Fischer-Tropsch technology has allowed Sasol in South Africa to produce liquid fuels commercially for decades using syngas derived from coal. The company uses the same basic technique today: syngas blown over a catalyst made of cobalt, nickel or iron transforms into various liquid hydrocarbons. Conveniently, the Fischer-Tropsch reaction gives off heat, and often this heat is

LARRY MAYER *Liaison International*

FLARED GAS heats the air uselessly around an oil well in Wyoming. Such supplies of natural gas are wasted because they are much more expensive than oil to transport by pipeline.



HARMFUL VEHICLE EMISSIONS were lowered somewhat in 1993, when U.S. regulations required that diesel fuel be reformulated to reduce pollution. Fuel derived from natural gas using Fischer-Tropsch synthesis creates even fewer emissions than reformulated diesel.

used to drive the oxygen compressors needed to make syngas.

Just which liquids emerge from the reaction depends on temperature. For example, running a reaction vessel at 330 to 350 degrees Celsius (626 to 662 degrees Fahrenheit) will primarily produce gasoline and olefins (building blocks often used to make plastics). A cooler (180 to 250 degree C) operation will make predominantly diesel and waxes. In any case, a mixture results, so a third and final step is required to refine the products of the reaction into usable fuels.

Refining synthetic crudes derived from gas is in many respects easier than working with natural crude oil. Synthetic crude contains virtually no sulfur and has smaller amounts of cancer-causing compounds than are found in conventional oil. So the final products are premium-quality fuels that emit fewer harmful substances.

A Partial Solution

This brute-force method of converting gas to liquids is reliable, but it is expensive because it uses so much energy. Conventional steam re-forming compresses methane and water vapor to about 30 times normal atmospheric pressure and heats these reactants to about 900 degrees C. And one must add more heat still, to coax the energy-hungry reaction continuously along. This extra heat comes from injecting a small amount of oxygen into the mixture, which combusts some of the methane (and, as an added benefit, makes more syngas). Chemists call this latter maneuver partial oxidation.

In general, syngas is generated using various combinations of steam re-forming and partial oxidation. In most cases, the process requires large quantities of oxygen—and oxygen is costly. Existing methods of separating oxygen from air rely on refrigeration to cool and liquefy it, an energy-intensive and expensive manipulation. Hence, lowering the cost of oxygen is the key to making syngas cheaply.

Fortunately, recent developments promise to revolutionize the way oxygen is produced over the next few years. One strategy is simply to work with air instead of pure oxygen. Syntroleum Corporation in Tulsa has developed a way to make liquid fuels using blown air and methane for the re-forming step, followed by Fischer-Tropsch synthesis. At sites where natural gas is sufficiently cheap (for example, places where it is now being flared), the process should prove profitable even at current crude oil prices. Together with Texaco and the English company Brown & Root, Syntroleum plans to build a commercial plant that will use this technique within two years.

Several other private companies, universities and government research laboratories are pursuing a wholly different approach to the oxygen problem: they are developing ceramic membranes through which only oxygen can pass. These membranes can then serve as filters to purify oxygen from air. Though still difficult and expensive to construct, laboratory versions work quite well. They should be commercially available within a decade.

Such materials could reduce the cost of making syngas by about 25 percent and lower the cost of producing liquid fuels by 15 percent. These savings would accrue because the production of syngas could be done at temperatures about 200 degrees lower than those currently used and because there would be no need to liquefy air. With cheap and plentiful oxygen, partial oxidation alone could supply syngas. This first step would then release energy rather than consume it.

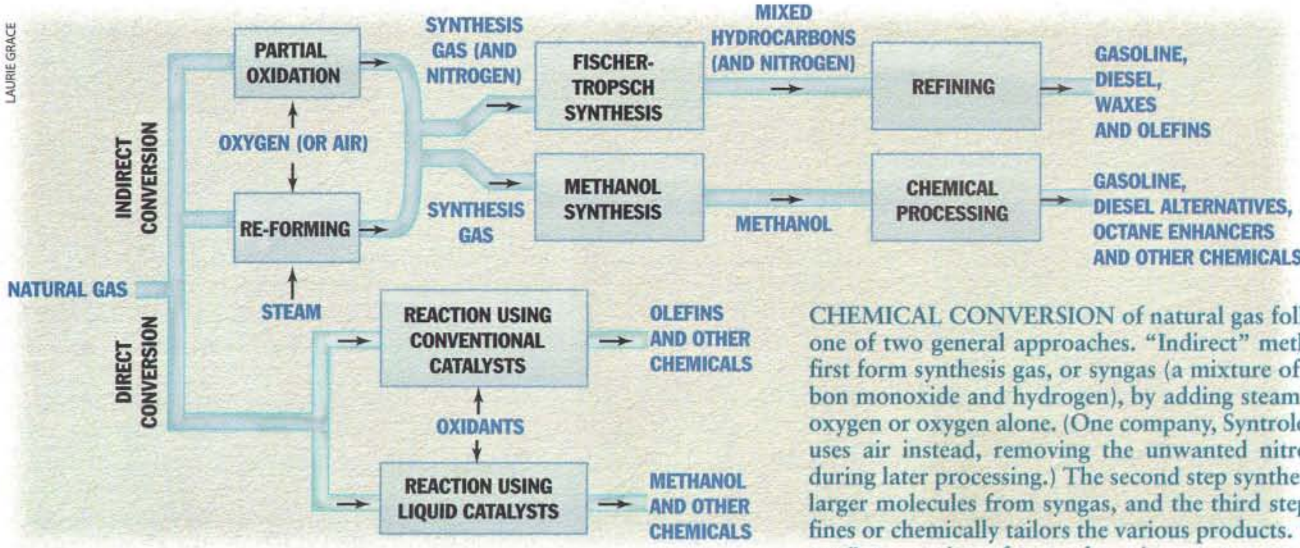
My Canadian colleagues and I, along with researchers at the University of Florida, are now attempting to create a different kind of ceramic membrane that would offer yet another advantage. The membranes we are trying to develop would remove hydrogen from the gas mixture, driving the partial oxidation of methane forward and providing a stream of pure hydrogen that could be used later in refining the final products or as an energy source itself.

We also expect to see significant improvements soon in the catalysts used to make syngas. In particular, researchers at the University of Oxford are studying metal carbides, and my colleagues at the Canadian Center for Mineral and Energy Technology are investigating large-pore zeolites. Both materials show great promise in reducing the soot generated during operation, a problem that not only plugs the reactor but also reduces the activity of the catalysts over time.

Cheaper than Oil?

Although the prospects for such brute-force methods of converting natural gas to liquid fuel improve every day, more ingenious techniques on the horizon would accomplish that transformation in a single step. This approach could potentially cut the cost of conversion in half, which would make liquid fuels produced from natural gas actually less expensive than similar products refined from crude oil.

Early efforts to achieve such "direct" conversion by using different catalysts and adding greater amounts of oxygen had produced mostly disappointment. The hydrocarbons that were formed proved more reactive than the methane supplied. In essence, they burned up



CHEMICAL CONVERSION of natural gas follows one of two general approaches. "Indirect" methods first form synthesis gas, or syngas (a mixture of carbon monoxide and hydrogen), by adding steam and oxygen or oxygen alone. (One company, Syntroleum, uses air instead, removing the unwanted nitrogen during later processing.) The second step synthesizes larger molecules from syngas, and the third step refines or chemically tailors the various products. "Direct" conversion of natural gas in one step requires an oxidant and may involve special liquid catalysts.

faster than they were produced. Unless the product is somehow removed from the reaction zone, yields are too low to be practical.

Fortunately, researchers have recently found ways to circumvent this problem. The trick is to run the reaction at comparatively mild temperatures using exotic catalysts or to stabilize the product chemically—or to do both. For example, chemists at Pennsylvania State University have converted methane to methanol directly using a so-called homogeneous catalyst, a liquid that is thoroughly mixed with the reactants and held at temperatures lower than 100 degrees C. And Catalytica, a company in Mountain View, Calif., has achieved yields for direct conversion that are as high as 70 percent using a similar scheme. Its liquid catalyst creates a relatively stable chemical intermediate, methyl ester, that is protected from oxidation. The final product (a methanol derivative) is easily generated with one subsequent step.

Methanol (also known as wood alcohol) is valuable because it can be readily converted to gasoline or to an octane-boosting additive. And in the near future methanol (either used directly or transformed first into hydrogen gas) could also serve to power fuel-cell vehicles on a wide scale. Thus, methanol can be regarded as a convenient currency for storing and transporting energy.

Moreover, the reactions used to synthesize methanol can be readily adjusted to churn out diesel alternatives such as dimethyl ether, which produces far fewer troublesome pollutants when it burns. So far dimethyl ether, like propane, has found little use as a transportation fuel because it is a gas at room temperature and pressure. But recently Air Products, a supplier of industrial gases in Allentown, Pa., announced the production of a dimethyl ether derivative that is liquid at ambient conditions. So this substitute for conventional diesel fuel would reduce emissions without major changes to vehicles and fueling stations.

Now You're Cooking with Gas

Scientists and engineers are pursuing many other possible ways to improve the conversion of natural gas into liquids. For instance, process developers are constantly improving the vessels for the Fischer-Tropsch reaction to provide better control of heat and mixing.

The most ambitious efforts now under way attempt to mimic the chemical reactions used by specialized bacteria that consume methane in the presence of oxygen to produce methanol. Low-temperature biological reactions of this kind are quite promising because they can produce specific chemicals using relatively little energy.

Whether or not this bold line of research ultimately succeeds, it is clear that even today natural gas can be converted into liquid fuels at prices that are only about 10 percent higher per barrel than crude oil. Modest improvements in technology, along with the improved economics that come from making specialty chemicals as well from gas, will broaden the exploitation of this abundant commodity in coming years. Such developments will also provide remarkably clean fuels—ones that can be easily blended with dirtier products refined from heavier crude oils to meet increasingly strict environmental standards. So the benefits to society will surely multiply as people come to realize that natural gas can do much more than just run the kitchen stove. SA

The Author

SAFAA A. FOUDA received a doctorate in chemical engineering from the University of Waterloo in 1976. Since 1981 she has worked at the CANMET Energy Technology Center, a Canadian government laboratory in Nepean, Ontario. There she manages a group of researchers studying natural gas conversion, emissions control, waste oil recycling and liquid fuels from renewable sources. Recently she headed an international industrial consortium intent on developing better methods to convert natural gas to liquid fuels.

Further Reading

METHANE CONVERSION BY OXIDATIVE PROCESSES: FUNDAMENTAL AND ENGINEERING ASPECTS. Edited by E. E. Wolf. Van Nostrand Reinhold, 1992.

THE DIFFERENT CATALYTIC ROUTES FOR METHANE VALORIZATION: AN ASSESSMENT OF PROCESSES FOR LIQUID FUELS. J. M. Fox III in *Catalysis Reviews: Science and Engineering,* Vol. 35, No. 2, pages 169–212; August 1993.

CONVERT NATURAL GAS INTO CLEAN TRANSPORTATION FUELS. M. Agee in *Hart's Fuel Technology and Management,* Vol. 7, No. 2, pages 69–72; March 1997.

REMOTE GAS STRATEGIES. World Wide Web site by Zeus Development Corporation available at http://www.remotegasstrategies.com/

THE AMATEUR SCIENTIST

by Shawn Carlson

The Pleasures of Pond Scum

If extraterrestrial scientists were ever to visit our planet, they might not pay much attention to relatively large, complex creatures like ourselves. Rather I suspect they would notice that the greatest diversity, and indeed most of the biomass, on the earth comes from its simplest residents—protozoans, fungi and algae, for example. By just about any objective measure, these organisms represent the dominant forms of life.

The world of algae is a particularly fascinating realm for amateur exploration. A single drop of pond water can harbor a breathtaking variety of microscopic species, with each algal cell constituting a complete plant. So if you want to view the essential biological underpinnings of all plant life, algae provide the best show in town.

Obviously, you will need a microscope to see the spectacle. An instrument with a magnification of about 120 times should be adequate. But the real secret to exploring this enthralling world successfully is to create small enclosures—"microponds," if you like—in which your specimens can flourish.

Canning jars filled with some water and soil make ideal environments. To let gases in and out, cut a hole one centimeter (about half an inch) in diameter in each lid and plug the opening with a wad of sterile cotton. You will also need a source of light for these diminutive plants. Because direct sunlight can rapidly warm your tiny ponds to temperatures that are lethal to algae, place them in a window with a northern exposure. Or set a full-spectrum bulb on a timer to provide a more controlled source of illumination.

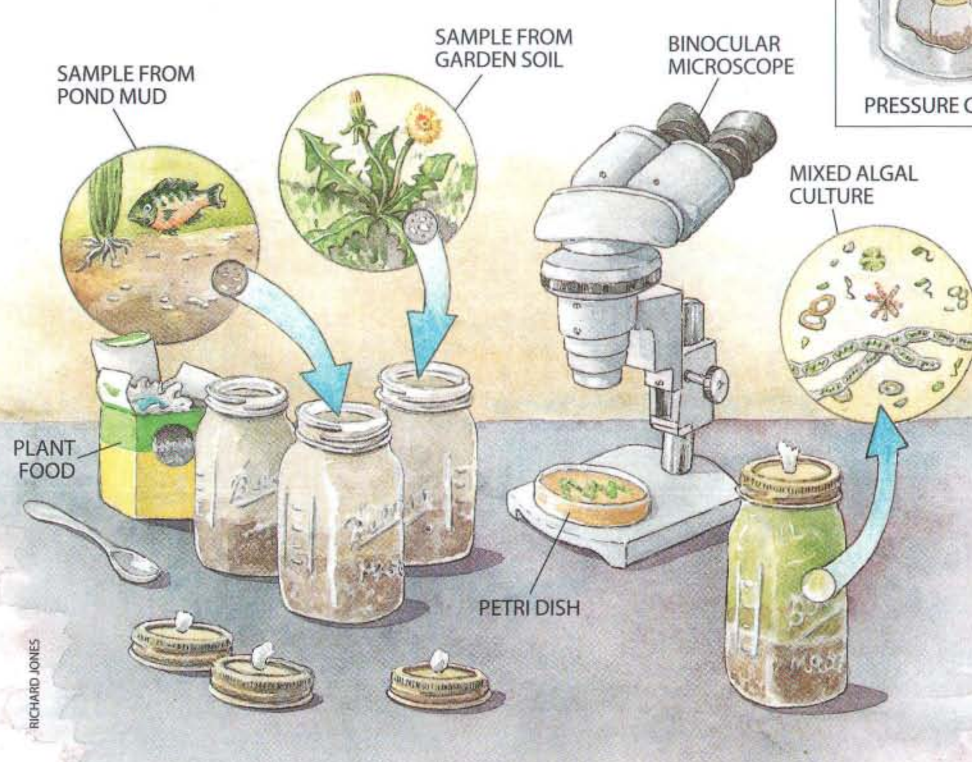
You must sterilize the soil, water, jar and lid before introducing any algae; otherwise stray bacteria could quickly take over. A special apparatus called an autoclave kills bacteria with heat under enough pressure to keep water from boiling. Scientists at a local university or commercial laboratory would probably be delighted to sterilize some materials for your home research using an autoclave. But a pressure cooker will also work.

You should first assemble everything. Add about four centimeters of dirt or mud and distilled water fortified with commercial plant food according to the manufacturer's instructions. Fill the jar to within two centimeters of the top, but do not screw the lids down yet. Place the prepared jars inside a plastic basin that is filled with two centimeters of water to prevent the base of the glass from getting too hot and breaking. Heat the jars for 20 minutes at 120 degrees Celsius (about 250 Fahrenheit).

If you don't have access to a pressure cooker, you can try vigorously boiling the water, jar and lid for at least 20 minutes in a covered pot and baking the soil in your oven at 180 degrees C for one hour. This method requires that you assemble the sterilized parts afterward, a procedure that risks contaminating your cultures with bacteria. If you do use this approach, let your containers sit for at least 10 days to make sure that your ponds are truly sterile; contaminating bacteria will turn the water cloudy as they multiply.

Otherwise, once the microponds have cooled completely, you should inoculate them with algae as soon as possible. A scraping from a piece of seaweed, a smidgen of pond mud, a pinch of garden soil, even a rubbing from the inside of a friend's aquarium are all great sources. Add such samples to a few jars and watch these aquatic gardens grow.

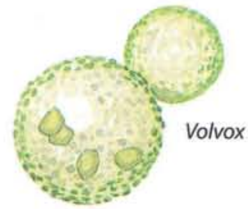
Your first cultures will probably contain a jungle of different single-celled plants. For scientific work, you will need to isolate individual species. You can do so by separating a small group of cells and implanting them in sterile agar, a gelatinous growth medium, which, incidentally, is itself made from algae. (A source of agar is listed at the end.)

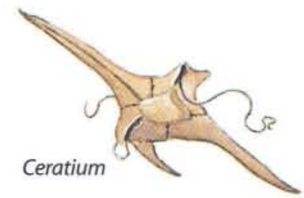


RICHARD JONES

***STERILIZED CANNING JARS** make ideal incubation chambers for growing algae taken from a garden or a pond.*

RICHARD JONES





This process is not as hard as it sounds. You begin by filling a set of sterilized petri dishes with agar to which you've added a small amount of plant food. Add the food just before setting the agar aside to gel. Next, use a sterile probe to take a tiny piece of the algae from one micropond. Swab the probe rapidly over the surface of the agar in a widely spaced zigzag pattern. Cells transferred onto the nutritious surface will take hold and grow in a few days into a splotchy garden of isolated groups plainly visible to the naked eye. Because cells will tend to shed in clumps from the original grab, most of the second-generation crop will be only slightly less diverse than their parent sample. But by taking a small amount from the center of one of these groups and then repeating the process on a second petri dish, you will create third-generation growths of even fewer species.

Some biologists recommend raising each succeeding generation by transferring a speck from a single petri dish to a sterilized culture jar, waiting for plants to multiply and then separating them again on a fresh agar surface. But I can sometimes get results more quickly by raising multiple generations on enriched agar, without culturing in jars as an intermediary step. Try this shortcut, but don't be surprised if you find you need to use the more laborious procedure.

It usually takes between three and five generations of swabbing and growing, but eventually your microponds will each contain a single algal species. Careful examination of samples under the microscope will reveal whether your cultures are pure. Once you have isolated several strains of algae in this way, the investigations you can make are limited only by your imagination. One useful class of experiments involves seeing how different chemicals affect growth.

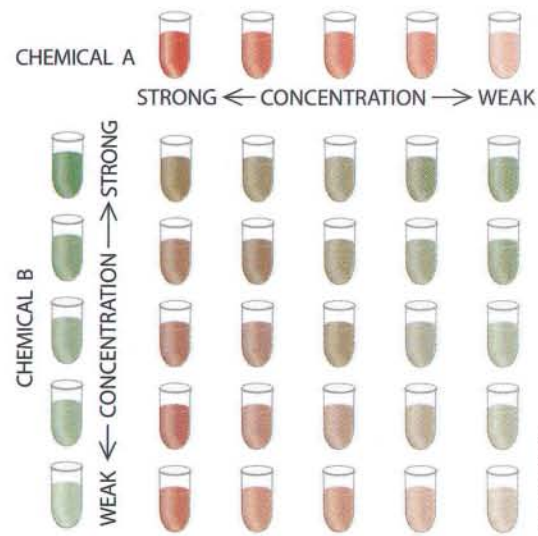
Your test results will be much simpler to compare if you begin each trial with samples containing roughly the same number of cells. One way to make up such samples is to dip blotting paper (sterilized with an autoclave or pressure cooker) into a dilute solution of sterile plant food and let it dry. Then soak the paper in molten beeswax for a few seconds until it is thoroughly impregnated. Remove the strips from the wax and return them to the solution of sterile plant food to harden. Transfer the waxed strips quickly to your culture jars. With luck, algal cells will grow uniformly over the treated paper. You can then cut out standardized samples using a sterile metal hole punch.

To start your experiment, set up a rack of 10 test tubes. Place a high concentration of the chemical you wish to evaluate in the first tube and then dilute each succeeding test tube by a factor of 10. That is, add nine parts of distilled water to one part of solution. (An eyedropper makes this task easier.) With 10 test tubes, you'll have a billion-fold difference in concentration between the first and last mixtures.

If you want to see whether the chemical in question improves growth, add three standard hole-punched samples and a fixed quantity of each of the dilutions to sterile culture jars. If you want to test whether the chemical can kill algae, you should place three standard samples for a prescribed time in each tube. Then use sterile tweezers to retrieve the samples and wash them gently with distilled water before placing them in the sterile culture jars.

Now monitor how the treated algae grow. Once you know roughly the minimum level at which the test chemical alters the amount of algae seen in the jar after one week, you can conduct a more tightly focused experiment to determine the critical concentration more precisely. A similar procedure lets you explore how algae respond to two separate chemicals [*see diagram above*]. This tactic is useful, for example, if you want to compare the effects of, say, a nitrogen-based fertilizer and a substance such as aspirin, which reduces the surface tension of water and so might make it easier for nutrients to get inside the cells.

This procedure will let you identify the optimal concentrations of two chemicals for growth of a particular species. You might also try to gauge how varying levels of heat and light affect the growth of algae, to learn their method of reproduction by microscopic examination or to identify some of their metabolic products using more sophisticated analytical apparatus (see, for example, the gas chromatograph described in the June 1966 Amateur Scientist). Remember, you can dive into this wondrous aquatic realm as deeply as you like. SA



JOHNNY JOHNSON

VARIABLE CONCENTRATIONS ***of two substances can be prepared by mixing equal amounts from separate dilutions of each chemical.***

For more information about this and other projects from the Amateur Scientist, consult the Society for Amateur Scientists World Wide Web page: www.thesphere.com/SAS/. Or you may write the society at 4735 Clairemont Square, Suite 179, San Diego, CA 92117, or call (619) 239-8807.

SOURCE

Powdered agar can be obtained from VWR Scientific Products; call (800) 727-4368 or (847) 459-6625. Ask for catalogue no. WLC 3049R.

MATHEMATICAL RECREATIONS

by Ian Stewart

Glass Klein Bottles

Alan Bennett is a glassblower who lives in Bedford, England. A few years ago he became intrigued by the mysterious shapes that arise in topology—Möbius bands, Klein bottles and the like—and came across a curious puzzle. A mathematician would have tried to solve it by doing calculations. Bennett solved it in glass. His series of remarkable objects, in effect a research project frozen in glass, is soon to become a permanent exhibit at the Science Museum in London.

Topologists study properties that remain unchanged even when a shape is stretched, twisted or otherwise distorted—the sole proviso being that the deformation must be continuous, so that the shape is not permanently torn or cut. (It is permissible to cut the shape temporarily, provided it is eventually reconnected so that points that were originally near one another across the cut end up near one another once again.) Topological properties include connectivity: Is the shape in one piece or several? Is the shape knotted or linked? Does it have holes in it?

The most familiar topological shapes appear at first sight to be little more than curious toys, but their implications run deep. There is the Möbius band, which you can make by taking a long strip of paper and gluing its ends together after giving the strip a twist. (Throughout this column, "twist" means "turn through 180 degrees," although sometimes this operation is known as a half-twist.) The Möbius band is the simplest surface that has only one side. If two painters tried to paint a giant Möbius band red on one side and blue on the other, they would eventually run into each other.

If you give the strip several twists, you get variations on the Möbius band. To a topologist, the important distinction is between an odd number of twists, which leads to a one-sided surface, and an even number, which leads to a two-sided surface. All odd numbers of twists yield surfaces that, intrinsically, are topologically the same as a Möbius band. To see why, just cut the strip, unwind all twists save one and join the cut up again. Because you removed an even number of twists, the cut edges rejoin into a simple Möbius band.

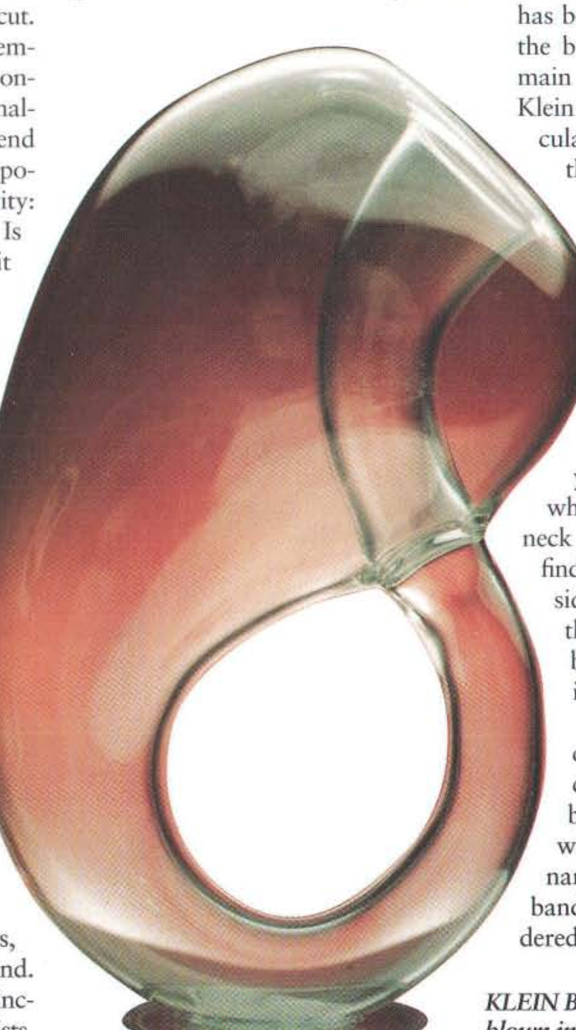
For similar reasons, all bands made with an even number of twists are topologically the same as an ordinary cylinder, which has no twists. The precise number of twists also has topological significance, however, because it affects how the band sits in its surrounding space. There are two different questions here—one about the intrinsic geometry of the band, the other about a band embedded in space. The first depends only on the parity (odd or even) of the number of twists; the second depends on the exact number.

The Möbius band has a boundary—those parts of the edge of the strip that don't get glued together. A sphere has no boundary. Can a one-sided surface have no edges at all? It turns out that the answer is yes, but no such surface can exist in three-dimensional space without crossing through itself.

This is no problem for topologists, who can imagine surfaces in space of higher dimensions or even in no surrounding space at all. For glassblowers, however, it is an unavoidable obstacle. The illustration below shows a glass Klein bottle blown by Bennett. Unlike an ordinary bottle, the "spout" or "neck" has been bent around, passed through the bottle's surface and joined to the main bottle from the inside. The glass Klein bottle meets itself in a small, circular curve; the topologist ignores that intersection when thinking about an ideal Klein bottle.

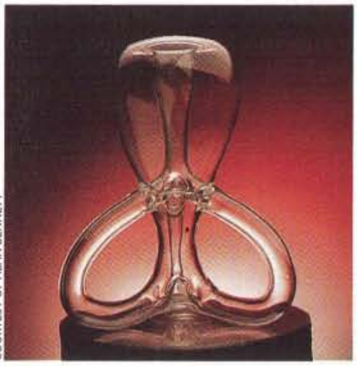
Imagine trying to paint a Klein bottle. You start on the "outside" of the large, bulbous part and work your way down the narrowing neck. When you cross the self-intersection, you have to pretend temporarily that it is not there, so you continue to follow the neck, which is now inside the bulb. As the neck opens up, to rejoin the bulb, you find that you are now painting the inside of the bulb! What appear to be the inside and outside of a Klein bottle connect together seamlessly: it is indeed one-sided.

Bennett had heard that if you cut a Klein bottle along a suitable curve, it falls apart into two Möbius bands. In fact, if you do this with a Klein bottle that sits in ordinary space like the glass one, those bands have a single twist. He wondered what kind of shape you had to

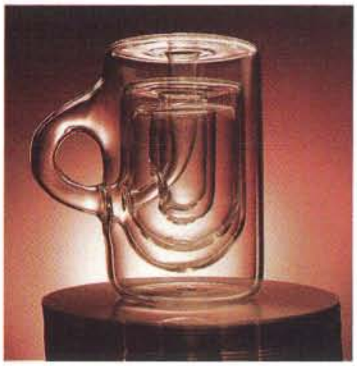


COURTESY OF ALAN BENNETT

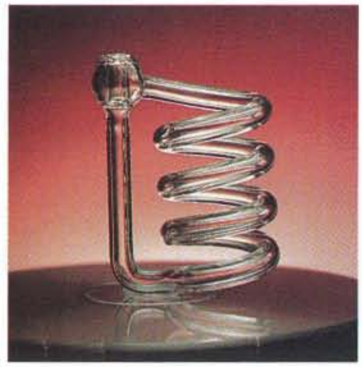
KLEIN BOTTLE, a one-sided surface, blown in glass by Alan Bennett



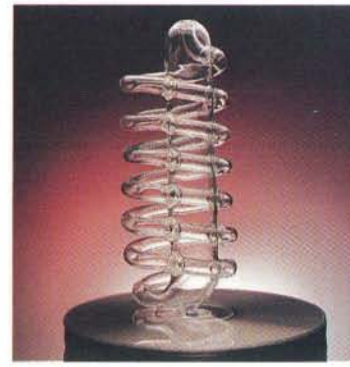
THREE-NECKED Klein bottle



NESTED SET of three Klein bottles



SPIRAL Klein bottle cuts into two seven-twist bands



VARIANT of spiral Klein bottle

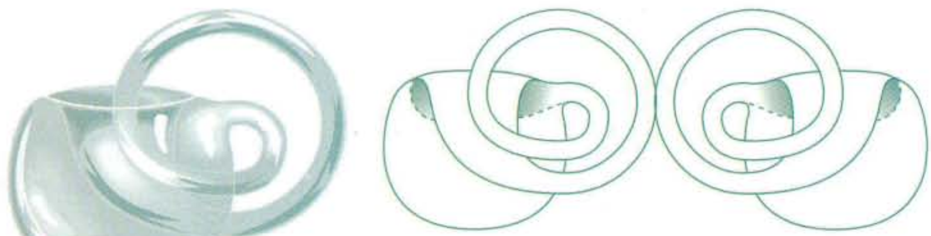
COURTESY OF ALAN BENNETT

cut up to get two three-twist Möbius bands. So he made many different shapes in glass and cut them up. He writes: "I find that if enough variations to the basic concept are made, or collected, the most logical or obvious solution to the problem usually becomes apparent."

Because he was looking for three-twist Möbius bands, Bennett tried all kinds of variations on the number three—such as bottles with three necks and, amazingly, sets of three bottles nested inside one another. He started to see, in his mind's eye, what would happen when they were cut up; he actually cut them up with a diamond saw to check.

The breakthrough was a very curious bottle whose neck looped around twice, forming three self-intersections. He named it the "Ouslam vessel," after a mythical bird that goes around in ever decreasing circles until it vanishes up its own rear end. If the Ouslam vessel is sliced vertically, through its plane of left-right symmetry—the plane of the paper in the drawing—then it falls apart into two three-twist Möbius bands. Problem solved.

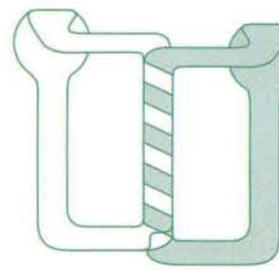
Like any mathematician, Bennett was now after bigger game. What about five-twist bands? Nineteen-twist bands? What was the general principle? Adding an extra loop, he quickly saw that five-twist bands would result. Every extra loop put in two more twists.



OUSLAM VESSEL, whose neck loops around twice, separates into two three-twist Möbius bands if sliced vertically. (The dotted lines are added as visual aids.)

Then he simplified the design, making it more robust, to produce spiral Klein bottles. The one depicted above, in the second photograph from the right, cuts into two seven-twist bands—and every spiral turn you add puts in two more twists. The photograph at the far right shows another variant on the same theme, a topological deformation of an ordinary spiral Klein bottle.

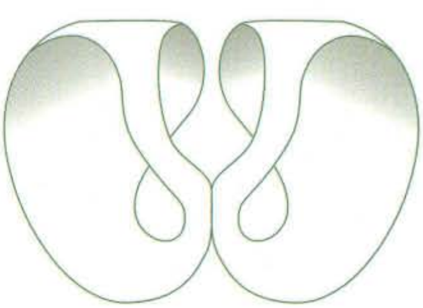
Having now seen the significance of spiral turns, Bennett realized that he could go back to the original Klein bottle by "untwisting" the spiral. The line along which the spiral Klein bottle should be cut would deform, too. As the spiral neck of the bottle untwisted, the cut line twisted up. So if you cut an ordinary Klein bottle along a spiral curve, you can get as many twists as you want—in this case, nine.



ORIGINAL Klein bottle cut along a spiral curve

ILLUSTRATIONS BY LAURIE GRACE

Now for a final curiosity. The original motivation for the work was the possibility of cutting a Klein bottle to get two one-twist Möbius bands. But you can also cut a Klein bottle along a different curve to get just one Möbius band. I'll leave you to work out how, and I'll provide Bennett's solution in a future Feedback. SA



TWO MÖBIUS BANDS result from cutting a Klein bottle along a curve.

FEEDBACK

Many readers sent information related to "Squaring the Square" [July 1997]. C. J. Bouwkamp of Eindhoven, an expert on squared squares, sent me a number of comments. First, I got some initials wrong: it should be A.J.W. Duijvestijn and C. J. Bouwkamp. Next, in 1969, P. J. Federico published simple perfect squared rectangles with one side double the other, and these were known even earlier. And, astonishingly, a cube of side 3,031,451 can be covered with 70 different squares. The breakdown is depicted at www.sciam.com —*I.S.*

REVIEWS AND COMMENTARIES

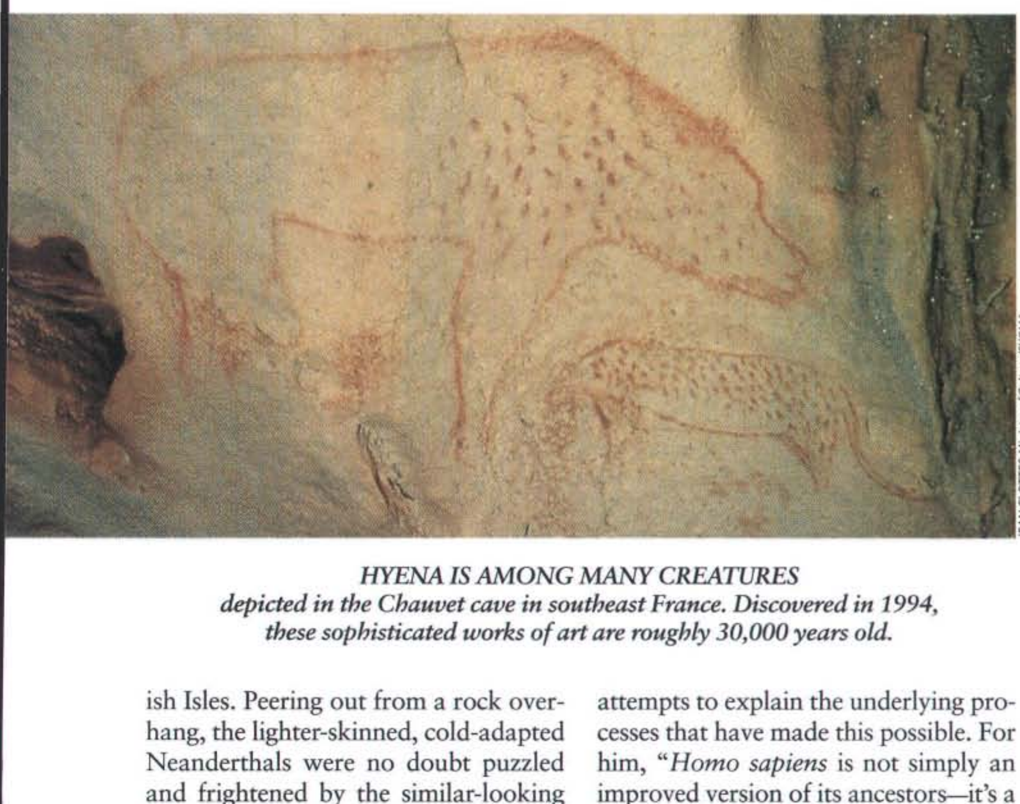
READING THE MINDS OF FOSSILS

Review by Donald Johanson

Becoming Human: Evolution and Human Uniqueness
BY IAN TATTERSALL
Harcourt Brace & Company, New York, 1998 ($27)

A momentous event transpired in human prehistory some 40,000 years ago when fully modern humans, armed with remarkably sophisticated tools and an unprecedented intelligence, began to populate Europe. Usually referred to as Cro-Magnons, these *Homo sapiens* encountered another species of humans, *H. neanderthalensis,* who had reigned unchallenged for perhaps 200,000 years throughout western Asia, most of Europe and even the British Isles. Peering out from a rock overhang, the lighter-skinned, cold-adapted Neanderthals were no doubt puzzled and frightened by the similar-looking but technologically superior beings that confronted them. These new humans brought with them a way of interacting with the world that initiated a slow but irreversible slide toward extinction for the Neanderthals.

In this superbly written book, Ian Tattersall combines his unique knowledge of the human fossil record, Paleolithic archaeology, primate behavior, prehistoric art, as well as the workings of the human brain and our extraordinary cognitive powers, to offer a convincing scenario of how we have come to hold dominion over the earth. Tattersall, who is chairman of the department of anthropology at the American Museum of Natural History in New York City, is one of our most competent and thoughtful chroniclers of human evolution, and here he ponders human uniqueness and attempts to explain the underlying processes that have made this possible. For him, "*Homo sapiens* is not simply an improved version of its ancestors—it's a new concept, qualitatively distinct from them in highly significant if limited respects." Trends in human evolution, such as increase in brain size, are discernible only in retrospect; such innovations were episodic and not the result of a process of hominization directed toward the emergence of ourselves.



JEAN CLOTTES *Ministry of Culture/SYGMA*

HYENA IS AMONG MANY CREATURES
depicted in the Chauvet cave in southeast France. Discovered in 1994, these sophisticated works of art are roughly 30,000 years old.

The "creative explosion" responsible for modern humans is perhaps most dramatically witnessed in Ice Age art. Examples of the art are found in caves such as Lascaux in France (and many, many others), which may have served as some sort of sanctuary, and in objects such as the Venus figurines, which even today, in the shape of soap bars, are bought by women as fertility aids. Tattersall believes that the art, supported by economic surpluses and executed by true artisans, was a reflection of how these early modern humans explained the world around them and their relation to that world.

Modern humans came to live in a world of their own creation; unlike Neanderthals, they were not simply reacting to "the world as nature presented it to them." Burials from the Upper Paleolithic (as the period from about 40,000 to 10,000 years ago is known) reflect a deeply spiritual side: these early *H. sapiens* offered elaborate grave goods to accompany the deceased in the afterlife. Neanderthal burials, on the other hand, are devoid of grave offerings and, like most of Neanderthal society, reflect an absence of ritual and symbolic activity.

Neanderthals were hunters, but their tools were far less sophisticated than those of early modern humans. They lacked hafted spears and relied on a Middle Paleolithic tool tradition, based on flake technology, that varied little over time and space. Upper Paleolithic blade technology, on the contrary, was truly innovative, changing over some 30,000 years with substantial elaboration of different tool types. This highly efficient use of raw materials demanded great skill and knowledge and resulted in long, thin blades with 10 times more cutting edge per lump of flint than those of Middle Paleolithic technology.

Technologically and symbolically superior *H. sapiens* brought Neanderthals to extinction some 30,000 years ago, when they drove the earlier humans into peripheral areas of Europe where resources were diminished or less familiar. Our success, Tattersall proposes, is largely the result of language, not simply the intuitive level of understanding and rudimentary communication characteristic of Neanderthals, but symbolic, syn-

tactic language, which distinguishes us even today. Language is fundamental to our ability to think; it is "more or less synonymous with symbolic thought," and human intellect is simply impossible in its absence. The rich archaeological record, and to some extent the anatomy of our vocal apparatus and enlargement of certain areas of the brain, supports the notion that Cro-Magnons were fully capable of language. Seeing modern humans as an abrupt departure from all that came before, Tattersall eliminates Neanderthals from our direct ancestry (contra the Multiregional theorists, who view Neanderthals and other archaic species of *Homo* as ancestral to modern humans; see "The Multiregional Evolution of Humans," by Alan G. Thorne and Milford H. Wolpoff; SCIENTIFIC AMERICAN, April 1992).

> *"At one cave in France, you have to paddle up an underground stream, then walk, wriggle, and crawl for two hours before finally, a mile underground, reaching the terminal chamber.... Today, entering such caves is an alien experience, even with the benefit of electric light, and you feel that [our forebears] must have been extraordinarily courageous to have undertaken such subterranean journeys by the faint and vulnerable light of fat lamps. But at one point..., far underground, there is a muddy ledge that bears the footprints of a child, no more than six years old, who had pranced down the ledge, digging in his or her toes—you can still see the imprints—just before the drop-off to the cave floor a foot below. At another site you can see where, a thousand yards underground, an adult must have taken the hand of a child and traced its finger across the soft surface of the cave wall. And at yet another, you can see where an adult had held a child's hand against the cave wall and blown pigment over it, leaving on the wall a negative imprint of the tiny fingers and palm."*
>
> —*from* Becoming Human

Many of Tattersall's inferences concerning human uniqueness are deeply founded in his theoretical approach to evolution: he embraces "punctuated equilibria" and shuns such modish formulations as Richard Dawkins's "selfish gene," E. O. Wilson's "sociobiology" and the ever popular "evolutionary psychology," which explains myriad human behaviors as genetically controlled leftovers from life in a hunting-gathering "ancestral environment." For Tattersall—and I totally concur—the process of evolution is not characterized by the gradual accumulation of small changes but by diversification and ultimately speciation. This perspective makes it easy to see how European populations of pre-Neanderthals became isolated, evolving into a distinct manifestation of our genus *Homo*—a classic speciation event. From this perspective, it seems logical, then, that we will not undergo any further speciation: our limitless cultural capacities permit cultural solutions to environmental changes (we build furnaces; we don't grow fur). And reproductive barriers simply no longer exist; air travel permits humans from around the globe to reproduce anywhere their plane ticket takes them.

By around 15,000 years ago, humans had undergone a diaspora that brought them to nearly every part of the planet. These humans, like their forebears 25,000 years earlier, were fundamentally the same as we are today. It is noteworthy that no matter where they lived, they were hunters (exceptional ones at that) and gatherers. Tattersall reminds us that, perhaps like the Pygmies of the Mbuti Forest in the Democratic Republic of the Congo, they saw themselves as part of the natural world, dependent on it and probably even respectful of it, like the Pygmies who cry, "Mother Forest, Father Forest."

All this was to change when we became agriculturists. In contrast to a hunting-gathering way of life that exploited but did not attempt to alter nature, the very essence of growing crops demanded modification of the natural environment. With this innovation began a battle with the natural world that has often led to extremely detrimental consequences, such as deforestation, soil degradation, aridification and so on.

But most important, it distanced us from nature and lessened our reverence for the natural world. We have become arrogant enough to believe that we are now outside of nature. In truth, however, as Tattersall poignantly reminds us, we will never escape our responsibilities to the global ecosystem. With this in mind, he concludes, "Barring disaster, we will almost certainly forever be the idiosyncratic, unfathomable and interesting creatures we have always been."

Let us hope that we live up to our name—*Homo sapiens*, "Man the Wise."

JEAN CLOTTES Ministry of Culture/SYGMA

***HORSES' HEADS* in the Chauvet cave**

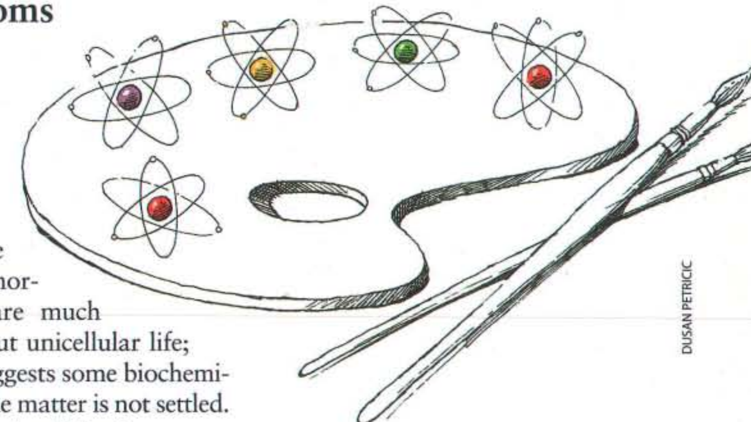
DONALD JOHANSON is professor of anthropology and director of the Institute of Human Origins at Arizona State University. His most recent book is From Lucy to Language, *written with Blake Edgar and published by Simon & Schuster in 1996.*

COMMENTARY

WONDERS

by Philip and Phylis Morrison

Painters' Atoms



Just west of New York City's Central Park, the great American Museum of Natural History prepares to move one of its finest mineral specimens. To receive this rarity, the builders are erecting a big concrete pillar, set deep into bedrock. No puny gemstone, this truck-size specimen is 15 tons of a near-stainless natural iron with a little nickel. Long, long ago that mass fell glowing out of the sky over Oregon. More than an unusual mineral, it is part of the history of asteroids and so an apt relic in its new home, the museum's planetarium.

Of course, iron metal itself abounds in Manhattan. Tens of millions of tons of steel, an iron-carbon alloy, form the bones of the skyline, as well as rail lines and bridge spans high and low. When melted from ore in a hot furnace, well-refined iron seeps and flows into the bottom of its cauldron, leaving a complex, rocky slag above it. Any cosmic ball that begins as a big enough jumble of iron-bearing rocks will do much the same under self-gravity. That close analogy was grasped very early and led to the conjecture of a structure for our earth that resembles our furnace contents. The remarkable tomography extracted over the past century from careful study of earthquake tremors has fully verified the presence of iron, both fluid and solid, at the earth's core, the magnetized ferrosphere.

Not all iron is buried deep. Certain bacteria concentrate iron oxides out of the long-stagnant water of peat bogs. Those dark sediments have for millennia supplied iron ores, more often for farmers' tools than for warriors' shields. Our industrial world remains bound to iron, although its dominance is under challenge. The ores smelted in modern ironworks are now mainly the amazing "banded iron formations" found in very old rocks around the world. Their red layers of ferric oxide, paper-thin inside a hard, quartzy matrix, can often be traced unbroken across 100 miles of the outcrop. Such enormous deposits are much older than any but unicellular life; their structure suggests some biochemical rhythm, but the matter is not settled.

In the 1960s the Athenians were proudly restoring their Parthenon, taking new marble from the very same mountain quarries that had furnished the stone when the temple was first built after victory over the Persians. Not all the results of the restoration were pleasing. The masterly workmanship was matched, but the new patches looked faintly gray-green, all but cold beside

> *Whereas gold ores are rare, iron oxides are ubiquitous. Why so much iron?*

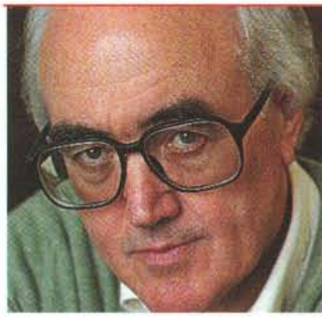
the delicate blush that so warmed the old columns. Only the newness can be blamed; greenish iron compounds, a minor ferrous impurity in all the white marble, had become red ferric oxide after 2,000 years of rains had activated the iron to link up with more oxygen.

Redness suffuses many vast landscapes, from the painted Arizona deserts to the flaming cliffs of the Gobi and even across space to the Sagan Station on the plains of Mars. Ferric oxide is reddest when full oxidation is followed by drying. Even before our species appeared, our hominid forebears used the vivid colorant red ocher as an ornament. The atoms of iron hold outer electrons unpaired in orbits far from filled, able to bond in various ways with oxygen, a plentiful atom that has electrons to spare. The common iron oxides and hydroxides include black magnetite, the yellow-brown umbers of limonite and the sharp red ocher of hematite. Complicated mixtures, often the presence of a few less common atomic species that resemble iron, extend the palette; crimson, magenta, orange and deep black add variety to our vistas, often even to much humbler reddish soils underfoot.

A more probing census is revealing. Such elements as iodine, silver, gold and lead are familiar enough yet are only nuances in the terrestrial recipe. Whereas gold ores are rare, iron oxides are ubiquitous. Chemistry and geology can sequester solid materials but cannot much affect the atomic count. Yet 10 times more atoms of iron are found on the earth than the total of all other atoms heavier than the iron group elements (the atoms of chromium, manganese, iron, cobalt and nickel). Why so much iron?

The atomic population in the sun and the stars is immature, nearly pristine; the list is strongly headed by the two primeval gases hydrogen and helium, containing the two simplest of atoms. Almost alone they endowed the very first stars. Over cosmic time, nuclear reactions in stellar interiors built basically all the rest of the atoms at fierce temperatures, a million times that of the sun's visible surface. There the typical photon is a gamma ray, not visible light, red or blue. In the radioactive emissions

Continued on page 87



COMMENTARY

CONNECTIONS
by James Burke

Turkish Delight

The new socket bayonet and flintlock musket turned warfare from hacking and yelling toward discipline and training.

I was watching TV one chilly evening recently and thinking about sun, sand and sea when suddenly on the screen there was an ad extolling the tourist and cultural attractions of one of my favorite holiday spots: Turkey. One side of the screen showed a tulip (almost the Turkish national emblem), and the other, the ruins of ancient Troy, the city that a German eccentric called Heinrich Schliemann "found" in 1873. A self-made businessman who had made a fortune in California's gold fields, Schliemann had gone on to market dyestuffs in Russia. Obsessed by the writings of Homer, he decided to spend a fortune trying to prove that the *Iliad,* Troy and all that poetic stuff about Helen launching a thousand ships had all really happened. (He failed, but his work would later stimulate real archaeologists to take a closer look.)

DUSAN PETRICIC

An egregious, bad-tempered man, Schliemann desperately needed an aura of scientific respectability, which he acquired from his partner in these efforts: a medical genius (and fellow Homer freak) named Rudolph C. Virchow, known as the "Pope of German Medicine." Virchow virtually kicked off public health in his day and is celebrated as the discoverer of cellular pathology. It was Virchow who made the momentous statement that was to change medicine: *Omnis cellula a cellula* ("All cells come from other cells"). By identifying the cell as the ultimate unit of life and disease, Virchow also helped to pave the way for chemotherapy 30 years later.

Virchow was also in Troy for his own reasons. He was an amateur anthropologist and interested in the history of human culture. Anthropology had been more or less invented in Germany by Johann F. Blumenbach, who, among other pursuits, related skull shape to racial classification. He did so by placing a skull between his feet and looking down at it. This became known among his adherents as the "Blumenbach position." Using it, Blumenbach divided humans into five racial groups, to which he gave names, one of which is still in relatively general use: "Caucasian."

Blumenbach investigated the case of the "Savage Boy," a child discovered in 1724 in Hannover, Germany, and said to be a living example of an early human. Blumenbach demolished this argument, but not before the boy had been sent to London where he was cared for (and exhibited and also much discussed among philosophers) by the queen's personal physician, Dr. John Arbuthnot, whose work on statistical probability galvanized a dull Dutchman by the name of William J. 'sGravesande.

In 1736 this person (whose life is described by even ardent biographers as "uneventful") was teaching Newtonian science at Leyden University and was visited by a Frenchman who was engaged in writing a general guide to that great English physicist's work. Perhaps because of advice from 'sGravesand, the Frenchman's book would make its author the most famous science writer in Europe. His name was Voltaire.

As it happens, Voltaire also knew Arbuthnot, because they had met when the French thinker was in London, at which time they had gone together to see the latest theatrical smash hit. This was John Gay's *The Beggar's Opera,* the first real lyric opera and a rumbustious satirical swipe at the political establishment. Apart from getting its author into deep doo-doo with the authorities, Gay's effort was given a boffo rating by the chattering classes and set box-office records by running for an unprecedented 62 performances. Gay's work had originally been talent-spotted by the manager of the theater in Lincoln's Inn Fields, John Rich, so when the reviews came out it was said that the extraordinary success of the piece would make "Gay rich and Rich gay."

Rich himself had an eye for what was going to catch on with theater groupies, which was why he also staged the first real ballet, by a fellow called John Weaver, who gleaned what he knew about dance steps from the recently published translation of a French book on choreography. The original author had in turn snitched the material from the dance master of Louis XIV, Pierre Beauchamp, who was the first to formalize the basic five feet positions and who developed a method for annotating dance movements, introducing French ballet terminology, like jeté and pas de deux.

Not surprisingly, as dance master to the king, Beauchamp worked closely with the music master to the king: Jean-Baptiste Lully, an Italian who had changed his name and who also wrote the first military band marches for the new French army. New it was because it was the first-ever standing (that is, full-time professional) army in Europe. This radical approach to military matters had been perfected by the French war minister, the marquis of Louvois, who realized that the megalomaniac maunderings of Louis XIV meant it was time to start making France great. This would involve doing things to make others less so. An army would help.

Louvois also saw how the new socket bayonet and flintlock musket combination was going to turn warfare away from hacking and yelling and more toward discipline and training, if the new weapons were to be used to their fullest advantage. That is, with clockwork precision, by uniform lines of men, maneuvering and firing by numbers. Louvois's new concept of a permanent army finally did away with the common practice of using mercenary troops to fight one's own battles.

This reengineering of events went over like a lead balloon with the traditional source of the best mercenaries in Europe: the Swiss. Principally the pike square technique had made them popular. This formation enabled large numbers of pikemen to protect small numbers of musketeers by standing round them with their 20-foot, steep-pointed pikes, which they would lower into a kind of hedgehog configuration anytime enemy cavalry appeared. At which point the cavalry would stop, and the musketeers would knock them off. The new musket and bayonet would do both jobs in one.

There was one bit of Switzerland that didn't care: the canton of Zurich, which, some time back, had already canceled its mercenary contracts, thanks to the activities of a Holy Roller–type named Huldrych Zwingli. By 1520 this firebrand religious reformer had effectively taken his community out of the Catholic Church with such un-Roman ideas as eating sausages during Lent, allowing priests to marry, taking down the statues of saints, conducting services in German instead of Latin and outlawing low-cut shoes. Some killjoy.

Zwingli's godson, Conrad Gesner, was an equally pious wimp, who tickled Huldrych's fancy with stuff like the Lord's Prayer in 22 languages and a giant catalogue of all books ever printed to that date. One other thing Gesner got up to—200 years ahead of Linnaeus—was formulating a new (the first) classification of animals based on their physiology and of plants (the first) based on their shapes and seeds. As part of this latter, botanical effort, in 1561 Gesner published a book containing the first European drawings of a knockout new flower, recently arrived from exotic foreign parts. The tulip on my TV screen the other night. SA

Wonders, continued from page 84

from Supernova 1987A, scientists detected the very gamma rays of iron-building, or at least of its embers, as the new matter exploded into space.

Iron of atomic mass 56 is some 90 percent of all our iron, and it has the distinction of having the most stable nucleus of all the elements. Smaller nuclei have too much surface, so their nucleons have too few neighbors for the best bonding. Heavier nuclei—themselves very rare—are less and less stable. They hold more and more positive electrical charges. By the rules of electrostatics, like charge repels like even across the entire nucleus, and any nucleus with more than 83 protons is unstable over geologic time.

To form the heavier nuclei of the earth, neutron capture and even nuclear fission are first requisites, not the simpler thermal exchange of a few nucleons. These heavier atoms are therefore very much rarer than atoms like iron, or most of the iron, that forms in the fierceness of star-core collapse. Any hotter furnace, and the nuclei themselves would break up. The iron group, with about 25 protons, give or take a few, has two special properties. Its atoms are the smallest to have electronic orbital complexity, and thus iron compounds display many colors under sunlight. Also, their nuclei strike the best possible balance between the attractive short-range nucleon-nucleon forces and the repulsive long-range electrostatic forces. That it is iron at the optimum, not one of its neighbors, is a detail that seems to have only a quantitative explanation.

We close this rather painterly story with one more remark. Red riverbeds are broad, but widespread, too, are the greens of forest and meadow. Their rich color is not dependent on any single metal atom but on a ring of atomic rings in chlorophyll a, within which carbons cooperate in shuffling electrons. The resulting states reflect out of all that white light only the yellow-green. Our own blood pigment, hemoglobin—a good color match to the ferric oxide of red ocher—has a central iron atom in a big complex, encircled by a ring of rings similar to that in chlorophyll. As far as we can tell, good physical chemists concur that the red of oxygenated blood is not direct from its iron atom but is a visual pun of molecular wit. SA

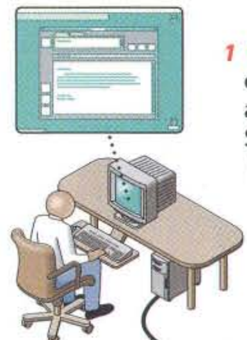
WORKING KNOWLEDGE

INTERNET ELECTRONIC MAIL

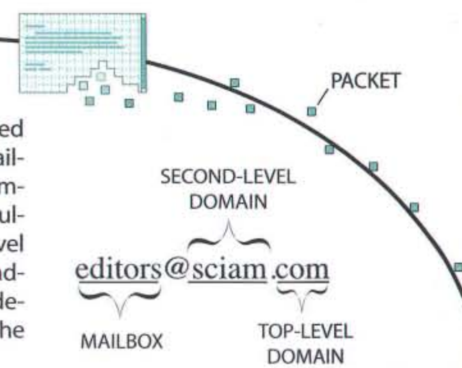
by Paul Hoffman
The Internet Mail Consortium

Billions of electronic mail (e-mail) messages move across the Internet every year. Sending electronic letters, pictures and data files, either across a building or across the globe, has grown so popular that it has started to replace some postal mail and telephone calls. This universal medium is no longer restricted to exchange of simple text messages and is now regularly used to deliver voice mail, facsimiles and documents that may include images, sound and video.

Typically, a message becomes available to the recipient within seconds after it is sent—one reason why Internet mail has transformed the way that we are able to communicate.

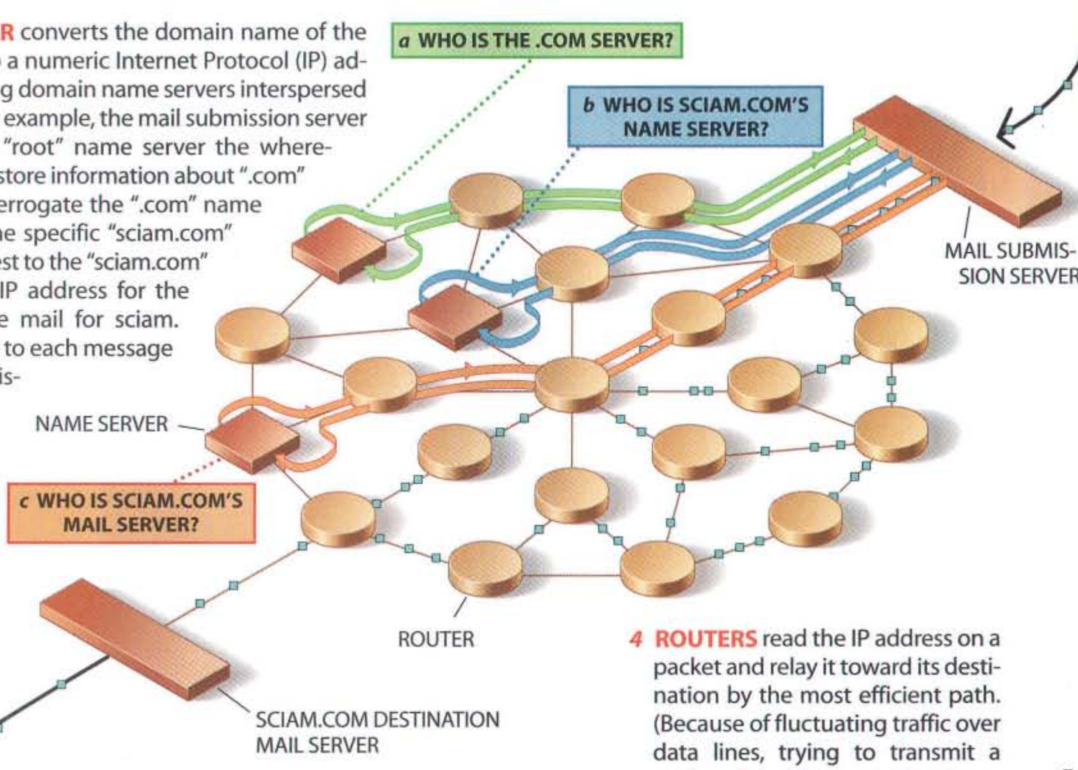


1 MESSAGE SENDER uses mail software, called a client, to compose a document, possibly including attachments such as tables, photographs or even a voice or video recording. System software, called Transmission Control Protocol (TCP), divides the message into packets and adds information about how each packet should be handled—for instance, in what order packets were transmitted from the sender. Packets are sent to a mail submission server, a computer on the internal network of a company or an Internet service provider.



2 INTERNET MAIL ADDRESSES attached to each message are in the form "mailbox@domainname"—one specific example being "editors@sciam.com." The multipart domain name denotes a top-level domain (".com") following the second-level domain ("sciam"). A message is delivered to an individual or a group by the mailbox name ("editors").

3 MAIL SUBMISSION SERVER converts the domain name of the recipient's mail address into a numeric Internet Protocol (IP) address. It does this by querying domain name servers interspersed throughout the Internet. For example, the mail submission server can first request from the "root" name server the whereabouts of other servers that store information about ".com" domains (*a*). It can then interrogate the ".com" name server for the location of the specific "sciam.com" name server (*b*). A final request to the "sciam.com" name server provides the IP address for the computer that receives the mail for sciam.com, which is then attached to each message packet (*c*). The mail submission server then transmits the packets to router machines dispersed throughout the Internet.



4 ROUTERS read the IP address on a packet and relay it toward its destination by the most efficient path. (Because of fluctuating traffic over data lines, trying to transmit a packet directly to its destination is not always the fastest way.) The packets of a single message may travel along different routes, shuttling through 10 or so routers before their journey's end.

5 DESTINATION MAIL SERVER places the packets in their original order, according to the instructions contained in each packet, and stores the message in the recipient's mailbox. The recipient's client software can then display the message.

JARED SCHNEIDMAN DESIGN